## Practical Machine Learning

## Workshop 2

Principal Components Analysis (PCA)
\& Support Vector Machine (SVM)

Dr. Suyong Eum

OSAKA UNVERSITY

## Principal Component Analysis (PCA)

## Principal Component Analysis (PCA): definition

A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

正

## Principal Component Analysis (PCA): definition

A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

$\left[\begin{array}{ll}8 & 1 \\ 1 & 8\end{array}\right]$

## Principal Component Analysis (PCA): definition

A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.


## Principal Component Analysis (PCA): definition

A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.


## Principal Component Analysis (PCA): definition

A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.


## Principal Component Analysis (PCA): intuition

$\square$ How to select a principal component?

- One that captures the largest variance of
$\square$ Why?
- Because we want to clearly see how each data point is related (close) each other.
- Then, which one (PC1 or PC2) is better?



> the data points.

## How to find the principal components showing the largest variance?

$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$

$\square$ Distance to data points from the mean along the axis of " $\mathrm{v}_{1}$ "
 $=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad$ variance $=4$

$$
\begin{array}{r}
=[-2 \sqrt{ } 2,-\sqrt{ } 2,0,0, \sqrt{ } 2,2 \sqrt{ } 2 \\
\operatorname{var}= \\
N-1
\end{array}
$$ -

## How to find the principal components showing the largest variance?

- Distance to data points from the mean along the axis of " $v_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $\mathrm{v}_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$



$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$

## How to find the principal components showing the largest variance?



$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$

- Distance to data points from the mean along the axis of " $v_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $\mathrm{v}_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$

$\operatorname{cov}(X)=\left[\begin{array}{ll}2.4 & 1.6 \\ 1.6 & 2.4\end{array}\right]$
(s)

## How to find the principal components showing the largest variance?

- Distance to data points from the mean along the axis of " $\mathrm{v}_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $\mathrm{v}_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$

$\square \operatorname{cov}(\mathrm{X})=\left[\begin{array}{ll}2.4 & 1.6 \\ 1.6 & 2.4\end{array}\right] \quad$ variance along the axis of " $x_{1}$ "


$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$

$$
9,0
$$

$$
\text { variance along the axis of " } x_{2} \text { " }
$$

## How to find the principal components showing the largest variance?



$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$



- $\operatorname{cov}(x)=V \Lambda V^{T}$
- Distance to data points from the mean along the axis of " $v_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $v_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$

- $\operatorname{cov}(\mathrm{X})=\left[\begin{array}{ll}2.4 & 1.6 \\ 1.6 & 2.4\end{array}\right] \quad \begin{aligned} & \text { variance along the axis of " } x_{1} \text { " } \\ & \text { variance along the axis of " } x_{2} \text { " }\end{aligned}$


## How to find the principal components showing the largest variance?


$X=\left[\begin{array}{l}x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6}\end{array}\right]=\left[\begin{array}{cc}-2 & -2 \\ -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ 2 & 2\end{array}\right]$
$\square \operatorname{cov}(x)=V \Lambda V^{T}$

- Distance to data points from the mean along the axis of " $\mathrm{v}_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $v_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$

$\square \operatorname{cov}(\mathrm{X})=\left[\begin{array}{ll}2.4 & 1.6 \\ 1.6 & 2.4\end{array}\right] \quad \begin{aligned} & \text { variance along the axis of " } x_{1} \text { " } \\ & \text { variance along the axis of " } x_{2} \text { " }\end{aligned}$

## How to find the principal components showing the largest variance?

- Distance to data points from the mean along the axis of " $\mathrm{v}_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $v_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$

$\square \operatorname{cov}(\mathrm{X})=\left[\begin{array}{ll}2.4 & 1.6 \\ 1.6 & 2.4\end{array}\right] \quad \begin{aligned} & \text { variance along the axis of " } x_{1} \text { " } \\ & \text { variance along the axis of " } x_{2} \text { " }\end{aligned}$

$$
\square \quad \operatorname{cov}(x)=V \Lambda V^{T}
$$

$\square \operatorname{cov}(\mathrm{x})=V \Lambda V^{T}$



$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$

## How to find the principal components showing the largest variance?



$$
X=\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x_{4} \\
x_{5} \\
x_{6}
\end{array}\right]=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
$$

- Distance to data points from the mean along the axis of " $\mathrm{v}_{1}$ "

$$
=[-2 \sqrt{2},-\sqrt{2}, 0,0, \sqrt{2}, 2 \sqrt{2}] \quad \text { variance }=4
$$

- Distance to data points from the mean along the axis of " $v_{2}$ "

$$
=[0,0,-\sqrt{2}, \sqrt{2}, 0,0] \quad \text { variance }=0.8
$$

$\square \operatorname{cov}(\mathrm{X})=\left[\begin{array}{ll}2.4 & 1.6 \\ 1.6 & 2.4\end{array}\right] \quad \begin{aligned} & \text { variance along the axis of " } x_{1} \text { " } \\ & \text { variance along the axis of " } x_{2} \text { " }\end{aligned}$

$$
\square \quad \operatorname{cov}(x)=V \Lambda V^{T}
$$



## How to find the principal components showing the largest variance?

1) Find the covariance matrix of data points.
```
>> x
x =
    -2 -2
    -1
    1
    -1 1
    1}
>> cov(x)
ans =
    2.4000 1.6000
    1.6000
```


## How to find the principal components showing the largest variance?

1) Find the covariance matrix of data points.
2) Obtain the eigen values and vectors of the covariance matrix: eigen decomposition.

| 2.4000 | 1.6000 |
| :--- | :--- |
| 1.6000 | 2.4000 |

```
\(1.6000 \quad 2.4000\)
4.00000



>> [vec, val] \(=\operatorname{eig}(\operatorname{cov}(x))\)
>> [vec, val] \(=\operatorname{eig}(\operatorname{cov}(x))\)
>> [vec, val] \(=\operatorname{eig}(\operatorname{cov}(x))\)
\(\mathrm{vec}=\)
\(\mathrm{vec}=\)
\(\mathrm{vec}=\)
    \(\begin{array}{rr}-0.70711 & 0.70711 \\ 0.70711 & 0.70711\end{array}\)
    \(\begin{array}{rr}-0.70711 & 0.70711 \\ 0.70711 & 0.70711\end{array}\)
    \(\begin{array}{rr}-0.70711 & 0.70711 \\ 0.70711 & 0.70711\end{array}\)
    \(\begin{array}{rr}-0.70711 & 0.70711 \\ 0.70711 & 0.70711\end{array}\)
    \(\begin{array}{rr}-0.70711 & 0.70711 \\ 0.70711 & 0.70711\end{array}\)
    \(\begin{array}{rr}-0.70711 & 0.70711 \\ 0.70711 & 0.70711\end{array}\)
val =
val =
val =
Diagonal Matrix
Diagonal Matrix
Diagonal Matrix
    \(\begin{array}{rr}0.80000 & 0 \\ 0 & 4.00000\end{array}\)
    \(\begin{array}{rr}0.80000 & 0 \\ 0 & 4.00000\end{array}\)
    \(\begin{array}{rr}0.80000 & 0 \\ 0 & 4.00000\end{array}\)
    0 4.00000
    0 4.00000
    0 4.00000

\section*{How to find the principal components showing the largest variance?}
1) Find the covariance matrix of data points.
2) Obtain the eigen values and vectors of the covariance matrix: eigen decomposition.
3) Sort the eigen vectors in descending order in terms of their corresponding eigen values.
- an eigen vector with the largest eigen value becomes the first principal component.


\(\gg\) vec, val] \(=\operatorname{eig}(\operatorname{cov}(x))\)
vec \(=\)
\(2^{\text {nd }}\) principal component


\section*{How to find the principal components showing the largest variance?}

Actually, there is a more convenient way of doing it.
\(\square\) It is called "Singular Value Decomposition" or SVD.
\[
\begin{aligned}
& \text { Eigen Value decomposition } \\
& \mathrm{X}^{\mathrm{T}} \mathrm{X}=\mathrm{V} \Lambda \mathrm{~V}^{\mathrm{T}}
\end{aligned}
\]
\begin{tabular}{|c|c|c|}
\hline \multicolumn{3}{|l|}{>> x} \\
\hline \multicolumn{3}{|l|}{\(\mathrm{x}=\)} \\
\hline \multicolumn{3}{|c|}{-2 -2} \\
\hline \multicolumn{3}{|c|}{-1 -1} \\
\hline \multicolumn{3}{|c|}{\(1-1\)} \\
\hline \multicolumn{3}{|c|}{-1 1} \\
\hline \multicolumn{3}{|l|}{1} \\
\hline \multicolumn{3}{|c|}{22} \\
\hline \multicolumn{3}{|l|}{\[
\begin{aligned}
& \gg \operatorname{cov}(x) \\
& \text { ans }=
\end{aligned}
\]} \\
\hline & 000 & 1.6000 \\
\hline & 000 & 2.4000 \\
\hline
\end{tabular}
>> [vec, val] = eig(cov(x))
>> [vec, val] = eig(cov(x))
vec =
vec =
    0.70711 0.70711
    0.70711 0.70711
    0.70711 0.70711
    0.70711 0.70711
val =
val =
Diagonal Matrix
Diagonal Matrix
    0.80000 0
    0.80000 0
\> [vec, val]=eig(transpose(x)*x)
\> [vec, val]=eig(transpose(x)*x)

\section*{How to find the principal components showing the largest variance?}

Eigen Value decomposition
\[
\mathrm{X}^{\mathrm{T}} \mathrm{X}=\mathrm{V} \Lambda \mathrm{~V}^{\mathrm{T}}
\]
\begin{tabular}{|c|c|c|}
\hline \multicolumn{3}{|l|}{>> x} \\
\hline \multicolumn{3}{|l|}{\(\mathrm{x}=\)} \\
\hline \multicolumn{3}{|c|}{\(\begin{array}{ll}-2 & -2\end{array}\)} \\
\hline \multicolumn{3}{|c|}{-1 -1} \\
\hline \multicolumn{3}{|c|}{\(1-1\)} \\
\hline \multicolumn{3}{|c|}{-1 1} \\
\hline \multicolumn{3}{|l|}{1} \\
\hline \multicolumn{3}{|c|}{22} \\
\hline \multicolumn{3}{|l|}{\[
\begin{aligned}
& \gg \operatorname{cov}(x) \\
& \text { ans }=
\end{aligned}
\]} \\
\hline & 4000 & 1.6000 \\
\hline & 600 & 2.4000 \\
\hline
\end{tabular}


\section*{How to find the principal components
\(\square\) Actually, there is a more convenient way of doing it.
It is called "Singular Value Decomposition" or SVD. \\ How to find the principal component
\(\square\) Actually, there is a more convenient way of doing it
\(\square\) It is called "Singular Value Decomposition" or SVD.} .

Singular Value Decomposition (SVD)
\[
\mathrm{X}=\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}
\]
\(\mathrm{X}=\mathrm{U} \Sigma \mathrm{V}^{\mathrm{T}}\)
\(\square\) Value Deco
 \(\square\)
\(\square\)

\(\square\)
\(\square\)

\(\square\)
\(\square\)
\(+2\)


\section*{How to find the principal components showing the largest variance?}

Eigen Value decomposition
\begin{tabular}{|c|c|}
\hline \multicolumn{2}{|l|}{>> x} \\
\hline \multicolumn{2}{|l|}{\(\mathrm{x}=\)} \\
\hline -2 -2 & \\
\hline -1 -1 & \\
\hline \(1-1\) & \\
\hline -1 1 & \\
\hline 11 & \\
\hline 22 & \\
\hline \[
\begin{aligned}
& \gg \operatorname{cov}(x) \\
& \text { ans }=
\end{aligned}
\] & \\
\hline 2.4000 & 1.6000 \\
\hline 1.6000 & 2.4000 \\
\hline
\end{tabular}
\[
\mathrm{X}^{\mathrm{T}} \mathrm{X}=\mathrm{V} \Lambda \mathrm{~V}^{\mathrm{T}}
\]
\begin{tabular}{|c|c|}
\hline \multicolumn{2}{|l|}{\(\gg\) [vec, val] \(=\operatorname{eig}(\operatorname{cov}(x))\) vec =} \\
\hline -0.70711 & 0.70711 \\
\hline 0.70711 & 0.70711 \\
\hline \multicolumn{2}{|l|}{val =} \\
\hline \multicolumn{2}{|l|}{Diagonal Matrix} \\
\hline 0.80000 & 0 \\
\hline 0 & 4.00000 \\
\hline
\end{tabular}
```

```
>> [vec, val]=eig(transpose(x)*x)
```

```
>> [vec, val]=eig(transpose(x)*x)
vec =
vec =
    0.70711 0.70711
    0.70711 0.70711
    0.70711 0.70711
    0.70711 0.70711
val =
val =
Diagonal Matrix
Diagonal Matrix
    4.0000
    4.0000
        020.0000
```

```
        020.0000
```

```
\[
\begin{array}{ll}
\text { vec }= \\
& \\
-0.70711 & 0.70711 \\
0.70711 & 0.70711 \\
\text { val }= \\
\text { Diagonal Matrix } \\
0.80000 & 0
\end{array}
\]

Singular Value Decomposition (SVD)
\[
\mathrm{X}=\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}
\]
\(X^{T} X=\left(U \Sigma V^{T}\right)^{T}\left(U \Sigma V^{T}\right)\)

\section*{How to find the principal components
\(\square\) Actually, there is a more convenient way of doing it.
It is called "Singular Value Decomposition" or SVD. \\ How to find the principal component
\(\square\) Actually, there is a more convenient way of doing it
\(\square\) It is called "Singular Value Decomposition" or SVD.}

\section*{How to find the principal components showing the largest variance?}
\begin{tabular}{|ll|}
\hline\(\gg\) [ven, val] \(=\operatorname{eig}(\operatorname{cov}(x))\) \\
vic \(=\) \\
-0.70711 & 0.70711 \\
0.70711 & 0.70711 \\
val \(=\) \\
Diagonal Matrix \\
0.80000 & 0 \\
0 & 4.00000 \\
\hline
\end{tabular}
```

```
>> [vec, val]=eig(transpose(x)*x)
```

```
>> [vec, val]=eig(transpose(x)*x)
vec =
vec =
    0.70711 0.70711
    0.70711 0.70711
    0.70711 0.70711
    0.70711 0.70711
val =
val =
Diagonal Matrix
Diagonal Matrix
    4.0000
    4.0000
        0 20.0000
```

        0 20.0000
    ```
```

    gonal Matrix
    ```
```

    gonal Matrix
    ```
\[
\begin{aligned}
\mathrm{X}^{\mathrm{T}} \mathrm{X} & =\left(\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}\right) \\
& =\mathrm{V} \Sigma^{\mathrm{T}} \mathrm{U}^{\mathrm{T}} \mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}} \\
& =\mathrm{V} \Sigma^{2} \mathrm{~V}^{\mathrm{T}}
\end{aligned}
\]

Singular Value Decomposition (SVD)
\[
\mathrm{X}=\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}
\]


T

Eigen Value decomposition
\[
\mathrm{X}^{\mathrm{T}} \mathrm{X}=\mathrm{V} \Lambda \mathrm{~V}^{\mathrm{T}}
\]

\[
2.4000
\]


\section*{How to find the principal components
\(\square\) Actually, there is a more convenient way of doing it.
It is called "Singular Value Decomposition" or SVD. \\ How to find the principal components
\(\square\) Actually, there is a more convenient way of doing it.
\(\square\) It is called "Singular Value Decomposition" or SVD.}



\section*{How to find the principal components showing the largest variance?}

\section*{Eigen decomposition}
\[
\mathrm{X}^{\mathrm{T}} \mathrm{X}=\mathrm{V} \Lambda \mathrm{~V}^{\mathrm{T}}
\]
\begin{tabular}{|l}
\begin{tabular}{l}
\(\gg\) \\
vec \(=\) \\
-0.70711
\end{tabular} \\
0.70711 \\
vec, val] \(=\operatorname{eig}(\operatorname{cov}(x))\) \\
val \(=\) \\
Diagonal Matrix \\
0.80000
\end{tabular}
\begin{tabular}{|ll|}
\hline\(\gg\) [vec, val]=eig(transpose \(\left.(x)^{*} x\right)\) \\
vec \(=\) \\
-0.70711 & 0.70711 \\
0.70711 & 0.70711 \\
val \(=\) \\
Diagonal Matrix \\
\multicolumn{2}{|c|}{} \\
4.0000 & 0 \\
0 & 20.0000
\end{tabular}
\begin{tabular}{|ll|}
\hline\(\gg\) [vec, val]=eig(transpose \(\left.(x)^{*} x\right)\) \\
vec \(=\) \\
-0.70711 & 0.70711 \\
0.70711 & 0.70711 \\
val \(=\) \\
Diagonal Matrix \\
4.0000 & 0 \\
0 & 20.0000
\end{tabular}

Singular Value Decomposition (SVD)
\[
\mathrm{X}=\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}
\]
\[
\begin{aligned}
X^{T} \mathrm{X} & =\left(\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}\right) \\
& =\mathrm{V} \Sigma^{\mathrm{T}} \mathrm{U}^{\mathrm{T}} U \Sigma \mathrm{~V}^{\mathrm{T}} \\
& =\mathrm{V} \Sigma^{2} V^{\mathrm{T}} \quad \Lambda=\Sigma^{2}
\end{aligned}
\]
\(\mathrm{X}=\mathrm{U} \Sigma \mathrm{V}^{\mathrm{T}}\)

Eigen value


(

\section*{How to find the principal components
\(\square\) Actually, there is a more convenient way of doing it.
\(\square\) It is called "Singular Value Decomposition" or SVD. \\ How to find the principal components
Actually, there is a more convenient way of doing it.
It is called "Singular Value Decomposition" or SVD.}

\section*{How to find the principal components showing the largest variance?}

Eigen decomposition
\(\mathrm{X}^{\mathrm{T}} \mathrm{X}=\mathrm{V} \Lambda \mathrm{V}^{\mathrm{T}}\)

Singular Value Decomposition (SVD)
\[
\mathrm{X}=\mathrm{U} \Sigma \mathrm{~V}^{\mathrm{T}}
\]
\begin{tabular}{|c|c|}
\hline & \(\chi^{\mathrm{T}} \mathbf{X}\) \\
\hline \multicolumn{2}{|l|}{> x} \\
\hline \multicolumn{2}{|l|}{\(\mathrm{x}=\)} \\
\hline \(\begin{array}{ll}-2 & -2\end{array}\) & \\
\hline -1 -1 & \\
\hline \(1 \begin{array}{ll}1 & -1\end{array}\) & \\
\hline -1 1 & \\
\hline 11 & \\
\hline 22 & \\
\hline \[
\begin{aligned}
& \gg \operatorname{cov}(x) \\
& \text { ans }=
\end{aligned}
\] & \\
\hline 2.4000 & 1.6000 \\
\hline 1.6000 & 2.4000 \\
\hline
\end{tabular}


\section*{Actually, there is a more convenient way of doing it. \\ \(\square\) It is called "Singular Value Decomposition" or SVD.}


\section*{Now we know how to find the principal components}
\[
\begin{aligned}
& \text { ( } \\
& X=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
\end{aligned}
\]

2 dimension data points can be represented into one dimension space \(\left(v_{1}\right)\)
\[
\left[\begin{array}{cc}
-2 \sqrt{2} & 0 \\
-\sqrt{2} & 0 \\
0 & 0 \\
0 & 0 \\
\sqrt{2} & 0 \\
2 \sqrt{2} & 0
\end{array}\right]
\]
\[
\mathbf{X}=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
\]

2 dimension data points can be represented into one dimension space \(\left(\mathrm{v}_{1}\right)\)

\(x_{1}\)
\[
\mathrm{v}_{2}=\left[\begin{array}{c}
-0.70711 \\
0.70711
\end{array}\right]
\]

\[
\mathrm{v}_{1}=\left[\begin{array}{l}
0.70711 \\
0.70711
\end{array}\right]
\]

\(x_{1}\)
\[
\mathbf{X}=\left[\begin{array}{cc}
-2 & -2 \\
-1 & -1 \\
1 & -1 \\
-1 & 1 \\
1 & 1 \\
2 & 2
\end{array}\right]
\]

2 dimension data points can be represented into one dimension space \(\left(\mathrm{v}_{1}\right)\)


Set the " \(\mathrm{v}_{2}\) " into zero
\(\mathrm{X} \_\)rot_zero \(=\mathrm{X} \_\)rot \(\cdot \mathrm{V}^{-1}\)

\[
\mathrm{X} \_ \text {rot }=\mathrm{X} \cdot \mathrm{~V}
\]

\[
\begin{aligned}
\mathrm{X}^{\prime} & =\mathrm{X} \_ \text {rot_zero } \cdot \mathrm{V}^{-1} \\
& =\mathrm{X} \_ \text {rot_zero } \cdot \mathrm{V}^{\mathrm{T}}
\end{aligned}
\]

\section*{Example}







2





 \\ \section*{\section*{2 +2 \\ \section*{\section*{2 +2 \\ \\ \(-\)} \\ \\ \(-\)}

\(\square\)
\(\square\)

\(\square\) \(\square\) T \(\square\)
 \(\square\) \(\square\) \(\square\)

 \(\square\) \(\square\) \(\square\) _

\section*{Example}

\(\mathrm{X}=\mathrm{U} \Sigma \mathrm{V}^{\mathrm{T}}\)
\begin{tabular}{|c|c|c|c|c|c|}
\hline \multicolumn{6}{|l|}{>> \([\mathrm{U}, \mathrm{S}, \mathrm{V}]=\operatorname{svd}(\mathrm{x})\)} \\
\hline \(\mathrm{U}=\) & & & & & \\
\hline -0.63246 & 0.00000 & 0.30819 & -0.30819 & 0.28637 & 0.57274 \\
\hline -0.31623 & -0.00000 & -0.63635 & 0.63635 & 0.13426 & 0.26851 \\
\hline 0.00000 & -0.70711 & 0.50000 & 0.50000 & 0.00000 & 0.00000 \\
\hline -0.00000 & 0.70711 & 0.50000 & 0.50000 & -0.00000 & -0.00000 \\
\hline 0.31623 & 0.00000 & -0.00399 & 0.00399 & 0.94140 & -0.11720 \\
\hline 0.63246 & 0.00000 & -0.00/99 & 0.00799 & -0.11720 & 0.76560 \\
\hline \multicolumn{6}{|l|}{\(\mathrm{s}=\)} \\
\hline \multicolumn{6}{|l|}{Diagonal Matrix} \\
\hline 4.4721 & 0 & & & & \\
\hline 0 & 2.0000 & & & & \\
\hline 0 & 0 & & & & \\
\hline 0 & 0 & & & & \\
\hline 0 & 0 & & & & \\
\hline 0 & 0 & & & & \\
\hline \multicolumn{6}{|l|}{\(\mathrm{v}=\)} \\
\hline 0.70711 & -0.70711 & & & & \\
\hline 0.70711 & 0.70711 & & & & \\
\hline
\end{tabular}

\section*{Example}


\section*{Example}
\[
\mathrm{X}^{\prime}=\mathrm{X} \_ \text {rot_zero } \cdot \mathrm{V}^{-1}=\mathrm{X} \_ \text {rot_zero } \cdot \mathrm{V}^{\mathrm{T}}
\]

\section*{How to use PCA for machine learning?}

A digit number with 64 dimension can be shown in 2 dimension space \(\left(v_{1}\right.\) and \(\left.v_{2}\right)\).

\[
\begin{array}{cccc}
v_{1} & v_{2} & & \\
{\left[\begin{array}{cccc}
a_{11} & a_{12} & \cdots & a_{1 m} \\
\vdots & \vdots & \vdots & \vdots \\
a_{n 1} & a_{n 2} & \cdots & a_{n m}
\end{array}\right]}
\end{array}
\]


(

\(\square\) Each digit number has 8 by \(8=64\) dimensions.
\(\square\) After SVD, the first two principal components are selected, and the data points with 64 dimension are plotted in two dimension.


\section*{Support Vector Machine (SVM)}



\(\qquad\) \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \section*{\\ \\ \section*{\\ \section*{\section*{(s) \\ \\ \\ \section*{\\ \\ \section*{\\ \section*{\section*{(s) \\ \\ \\ \section*{\\ \\ \section*{\\ \section*{\section*{(s) \\ \\ \\ \section*{\\ \\ \section*{\\ \section*{\section*{(s) \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\  \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ ( \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\ \(\qquad\) \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\ \footnotetext{
 \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\ 
} \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\      \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ (a) \\ \\ \\ \\ (a) \\ \\ \\ \\ (a) \\ \\ \\ \\ (a) \\ \\ \\  \\ \\ \\  \\ \\ \\  \\ \\ \\ \\ \\ \\ 4 \\ \\ \\ \\ \\ \\ 4 \\ \\ \\ \\ \\ \\ 4 \\ \\ \\ \\ \\ \\ 4 \\ \\ \\ \\ \\ \\ 4 \\ \\ \\ \\ \\ \\ 4 \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\ (asses) \\ \\ \\ \\ (asses) \\ \\ \\ \\ (asses) \\ \\ \\ \\ (asses) \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\  \\ \\ \\ \\ ( \\ \\ \\ \\ ( \\ \\ \\ \\ ( \\ \\ \\ \\ ( \\ \\ \\ \\ \\ \\ ? \\ \\ \\ \\ \\ \\ ? \\ \\ \\ \\ \\ \\ ? \\ \\ \\ \\ \\ \\ ? \\ \\ \\ \\ \\ \\ ? \\ \\ \\ \\ \\ \\ ? \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\  \\ \\ \\ \\ \\ \\ (} \\ \\ \\ \\ \\ \\ (} \\ \\ \\ \\ \\ \\ (} \\ \\ \\ \\ \\ \\ (} \\ \\ \\ \\ \\ \\ (} \\ \\ \\ \\ \\ \\ (}

\section*{Terminology used in this lecture}

\section*{}
 ．

教 O －  O





Support
Vectors －






\begin{abstract}

\end{abstract}

E
\(+\) －

\section*{³²argin distance}

\(\mathrm{X}^{c}=\mathrm{X}^{b}+\|r\| \frac{\mathrm{W}}{\|\mathrm{W}\|}\)
\[
\mathrm{x}^{c}=\mathrm{x}^{b}+\|r\| \frac{\mathrm{w}}{\|\mathrm{w}\|}
\]
\[
\boldsymbol{X}_{2} \quad w_{2} x_{2}+w_{1} x_{1}+w_{0}=y(\mathrm{x})
\]


\(\frac{X_{2}}{\mathcal{N} \|} \quad w_{2} x_{2}+w_{1} x_{1}+w_{0}=y(\mathrm{x})\)
\(\frac{X_{2}}{\mathcal{N} \|} \quad w_{2} x_{2}+w_{1} x_{1}+w_{0}=y(\mathrm{x})\)
.
.

號
\[
x_{1}
\]
\(\mathrm{X}^{c}=\mathrm{X}^{b}+\|r\| \frac{\mathrm{W}}{\|\mathrm{W}\|}\)
\(\mathrm{X}^{c}=\mathrm{X}^{b}+\|r\| \frac{\mathrm{W}}{\|\mathrm{W}\|}\)

\(\qquad\)

Ma
.

\footnotetext{
－
}

\(\mathbf{W} \quad \begin{gathered}\text { Unit vector showing } \\ \text { the direction only }\end{gathered}\)
\(\mathbf{W} \quad \begin{gathered}\text { Unit vector showing } \\ \text { the direction only }\end{gathered}\)

\(\mathbf{W} \quad \begin{gathered}\text { Unit vector showing } \\ \text { the direction only }\end{gathered}\)
\(\mathbf{W} \quad \begin{gathered}\text { Unit vector showing } \\ \text { the direction only }\end{gathered}\)
\[
\| \frac{\mathrm{W} \text { Unit vector showing }}{\|\mathrm{W}\|} \text { the direction only }
\]

Unit vector showing
the direction only
Unit vector showing
the direction only

\(\mathbf{X}^{c} \mathbf{X}^{b} \mathbf{W} \quad\) Unit ie
\({ }_{x}{ }^{b} b \quad \mathbf{W}\) Unit vt
\(\mathbf{X}^{c} \mathbf{X}^{b} \mathbf{W} \quad\) Unit vt
\(\mathbf{X}^{c} \mathbf{X}^{b} \mathbf{W} \quad\) Unit ie
\({ }_{x}{ }^{b} b \quad \mathbf{W}\) Unit ie
\({ }_{x}{ }^{b} b \quad \mathbf{W}\) Unit vt

\section*{Margin distance}
\[
\mathrm{X}^{c}=\mathrm{X}^{b}+\|r\| \frac{\mathrm{W}}{\|\mathrm{~W}\|} \text { (size of the vector } \begin{aligned}
& \text { Unit vector showing } \\
& \text { the direction only }
\end{aligned}
\]

Let＇s multiply \(w^{\top}\) and add \(w_{0}\) in both sides．
\(w^{\top} x^{c}+w_{0}=w^{\top} x^{b}+w_{0}+w^{\top}\|r\| \frac{w}{\|w\|}\)
40
\[
w_{2} x_{2}+w_{1} x_{1}+w_{0}=y(\mathrm{x})
\]
\[
\begin{aligned}
& w^{T} x^{c}+w_{0}=y\left(x^{c}\right) \\
& W^{T} x^{b}+w_{0}=0
\end{aligned}
\]



\section*{}
，

\(r \| \frac{\mathbf{W}}{\|\mathbf{W}\|} \begin{gathered}\text { Unit vector showing } \\ \text { the direction only }\end{gathered}\)

\[
\mathrm{w}^{\mathrm{T}} \mathrm{X}^{c}+\mathrm{w}_{0}=\mathrm{w}^{\mathrm{T}} \mathrm{X}^{b}+\mathrm{w}_{0}+\mathrm{w}^{\mathrm{T}}\|r\| \frac{\mathrm{w}}{\|\mathrm{w}\|}
\]
\[
\mathbf{w}^{\mathrm{T}} \mathbf{x}^{c}+\mathbf{w}_{0}=\mathbf{w}^{\mathrm{T}} \mathbf{x}^{b}+\mathbf{w}_{0}+\mathbf{w}^{\mathrm{T}}\|r\| \frac{\mathbf{W}}{\|\mathbf{W}\|}
\]

 Margin distance
\(\mathbf{x}^{c}=\mathrm{X}^{b}+\|r\|\)
\(\begin{gathered}\text { Size of the } \\ \left(x^{b}->x^{c}\right)\end{gathered}\)
\(\mathbf{W}^{\mathrm{T}} \mathrm{X}^{c}+\mathrm{w}_{0}=\)
\(=\)
\(\square\)都

促
\[
40
\]

\(\square\)
\(\square\)
\(W^{T} x^{c}+w^{\prime}\)
\(>\)
\(\qquad\)


\(X^{c}=X^{\left(x^{b} \rightarrow x^{c}\right)}\)
－

> 蹅

\(\square L^{T} t^{T} s\) multiply \(W^{T}\) and add \(W_{0}\) in both sides．


解


\section*{Margin distance}


．

\(\qquad\) ．
（

\(\square\)

\footnotetext{
\(\qquad\)
}

2
\[
\mathrm{X}^{c}=\mathrm{X}^{b}+\|r\| \frac{\mathrm{W}}{\|\mathrm{~W}\|} \text { Unit vector showing } \begin{gathered}
\text { Size of the vector } \\
\left(\mathrm{x}^{b}-\mathrm{x}^{c}\right) \\
\text { the direction only }
\end{gathered}
\]

Let＇s multiply \(\mathrm{w}^{\top}\) and add \(\mathrm{w}_{0}\) in both sides．
\[
\begin{aligned}
& \mathrm{w}^{\mathrm{T}} \mathrm{x}^{c}+\mathrm{w}_{0}=\mathrm{w}^{\mathrm{T}} \mathrm{x}^{b}+w_{0}+\mathrm{w}^{\mathrm{T}}\|r\| \frac{\mathrm{w}}{\|\mathrm{w}\|} \\
& \mathrm{y}\left(\mathrm{x}^{\mathrm{c}}\right)=\mathrm{w}^{\mathrm{T}}\|r\| \frac{\mathrm{w}}{\|\mathrm{w}\|}
\end{aligned}
\]

\(\qquad\)
\[
\xrightarrow{\text { cos }\|r\|}
\]
\(X_{1}\)

\[
w_{2} x_{2}+w_{1} x_{1}+w_{0}=y(\mathrm{x})
\]
\[
\begin{aligned}
& \mathrm{w}^{\mathrm{T}} \mathrm{x}^{c}+w_{0}=y\left(\mathrm{x}^{\mathrm{c}}\right) \\
& \mathrm{w}^{\mathrm{T}} \mathrm{x}^{b}+\mathrm{w}_{0}=0
\end{aligned}
\]
\[
\begin{aligned}
& \mathrm{w}^{\mathrm{T}} \mathrm{x}^{c}+w_{0}=y\left(\mathrm{x}^{\mathrm{c}}\right) \\
& \mathrm{w}^{\mathrm{T}} \mathrm{x}^{b}+w_{0}=0
\end{aligned}
\]

\section*{Margin distance}
\[
\left.\mathrm{x}^{c}=\mathrm{x}^{b}+\|r\| \frac{\mathrm{W}}{\|\mathrm{~W}\|} \text { (xiz-2x}\right) \text { tUne vector direction only }
\]
\(\square\) Let's multiply \(\mathrm{w}^{\top}\) and add \(\mathrm{w}_{0}\) in both sides.
\[
\begin{aligned}
& \frac{\mathbf{W}}{\mathbf{W}} \|
\end{aligned}
\]
\[
\begin{aligned}
& \mathrm{w}^{\mathrm{T}} \mathrm{x}^{c}+\mathrm{w}_{0}=\mathrm{w}^{\mathrm{T}} \mathrm{x}^{b}+\mathrm{w}_{0}+\mathrm{w}^{\mathrm{T}}\|r\| \frac{\mathrm{w}}{\|\mathrm{w}\|} \\
& \mathrm{y}\left(\mathrm{x}^{\mathrm{c}}\right)=\mathrm{w}^{\mathrm{T}}\|r\| \frac{\mathrm{w}}{\|\mathrm{w}\|} \\
& \|r\|=\frac{\mathrm{y}\left(\mathrm{x}^{\mathrm{c}}\right)}{\|\mathrm{w}\|}
\end{aligned}
\]
.


\begin{abstract}

\end{abstract} --



\section*{Margin distance \\ Marg}
 ，
\[
\mathbf{X}^{c}=\mathrm{X}^{b}+\|r\| \frac{\mathbf{W}}{\|\mathbf{W}\|} \quad \begin{gathered}
\text { Size of the vector } \\
\left(\mathrm{x}^{\left.\mathrm{b}->\mathrm{x}^{c}\right)}\right. \\
\text { the direction only }
\end{gathered}
\]
 －
正
\(\square\) Finding a decision boundary which maximizes


\section*{roble formulation}
\(\qquad\)



\(X\)
W
\[
x^{b-}\|r\|
\]

\(x_{2}\)


都

\(\square\)
\[
\begin{gather*}
\\
\hline
\end{gather*}
\]
\[
\max \|r\|=\frac{1}{\|\mathrm{w}\|}
\]

-



1

\(\qquad\)

2
\(\qquad\)
\(\qquad\)
\(\qquad\)
\(\qquad\)
\[
\begin{align*}
& \text { Finding a decision boundary which ma) } \\
& \text { the margin. } \\
& \max \|r\|=\frac{1}{\|\mathrm{~W}\|} \\
& \text { s.t. } \\
& t_{n} y\left(\mathrm{x}_{n}\right)>0 \quad \begin{array}{l}
\text { Every data points are } \\
\text { classified correctly. }
\end{array} \\
& \begin{cases}t_{n}=+1, & y\left(\mathrm{x}_{n}\right)>0 \\
t_{n}=-1, & y\left(\mathrm{x}_{n}\right)<0\end{cases}
\end{align*}
\]
the margin.
max \(\|r\|=\frac{1}{\|w\|}{ }^{\text {s.t. }}\)
\(t_{n} y\left(\mathrm{X}_{n}\right)>0\)
\(\left\{\begin{array}{l}t_{n}=+1, \\ t_{n}=-1, \\ 45\end{array}\right.\)
\(y\left(x_{n}\right)>0\)
the margin.
max \(\|r\|=\frac{1}{\|\mathrm{w}\|}\)
s.t.
\(t_{n} y\left(\mathrm{X}_{n}\right)>0\)
\(\left\{\begin{array}{l}t_{n}=+1, \\ t_{n}=-1, \\ 45\end{array}\right.\)
\(y\left(\mathrm{x}_{n}\right)>0\)
\[
r_{1}
\]
■
\[
w_{2} x_{2}+w_{1} x_{1}+w_{0}=0
\]

都
\[
-
\]


\footnotetext{
\title{
Led er \\ r of
}
\(-\)
}
\[
\begin{cases}t_{n}=+1, & y\left(\mathrm{x}_{n}\right)>0 \\ t_{n}=-1, & y\left(\mathrm{x}_{n}\right)<0\end{cases}
\]
\(\qquad\)

oundary which maximizes
.

\footnotetext{


\title{
 \\ \(\qquad\) \\  \\ .
}
}

\section*{47ow about non-linearly separable case? \\ }
 ?





 T +
\(\qquad\)

\(\qquad\)




\[
0
\]







\(\qquad\)
\(\qquad\)







\(\square\)

 -
\(\square\)


號

\(\qquad\)


\(\mathrm{X}_{n}+\mathrm{W}_{0} \geq\),
\(\left.\mathrm{x}_{n}+\mathbf{w}_{0}\right) \geq 1, \quad \forall n\)
\(\left.\mathrm{x}_{n}+\mathrm{w}_{0}\right) \geq 1, \quad \forall n\)
\(\xrightarrow[\Delta]{\sim}\)
\(\xrightarrow[\Delta]{\sim}\)
\(\xrightarrow[\Delta]{\sim}\)
\(\left.+w_{0}\right) \geq 1, \quad \forall n\)
-



\[
\begin{equation*}
1 \tag{1}
\end{equation*}
\]

\(\left.\mathbf{W}_{0}\right) \geq 1, \quad \nabla n\)



CA







2

\section*{ \\ \\ \\ 地 \\ \\ \\ 地 \\ \\ \\ 地 \\ \\  \\ \\  \\  \\  \\  \\ － \\ }

\section*{Soft margin SVM}

\section*{50ption 1：soft margin SVM
}


\(\qquad\)
\[
\begin{gathered}
\text { s. slack } \\
\mathrm{w}^{\mathrm{T}} \mathrm{X}+w_{0}=1 \\
\mathrm{w}^{\mathrm{T}} \mathrm{X}+w_{0}=0
\end{gathered}
\]
保
        R


1
1
5

1
1
5
5

1
1
1
1
5
\(\begin{array}{ll} \\ 10 & w_{n}\left(w^{T} X_{n}+w_{0}\right) \geq 1,\end{array}\)
            \(1 \times 1\)
            \(1 \times 1\)
5
50

1
1

\(\qquad\)
\(-1\)五

\(\square\)



D
－

\begin{abstract}

\end{abstract}


 －

\(-1\)
\(\square\)
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
\(\square\)
\(\square\)

者


 \(+\)

```

                                    *
    ```
                                    *
```

                                    *
                                    \square
    ```
                                    \square
```

                                    \square
    ```
\(\square\)
Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\] the constraint：
\[
\begin{gathered}
\mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=1 \\
\mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=0 \\
\mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=-1
\end{gathered}
\]
\(t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n\)
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\] 
  （ \(\qquad\)

\section*{53ption 1：soft margin SVM \\ }
－For the data points which are non－separable，we relax

 \(\square\) 五



\[
0
\]

\begin{abstract}
．
\end{abstract} ．
\(\qquad\) \(\square\)
\(\square\)
\(\square\)


 
                            Remember the constraint below?
\(t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n\)\(\quad \begin{aligned} & \text { For the data points which are non -separable, we relax } \\ & \text { the constraint: } \\ & t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n}^{3} \quad \forall n \quad \varepsilon_{n} \geq 0\end{aligned}\)
R

\section*{50ption 1：soft margin SVM}
\(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]

For the data points which are non－separable，we relax the constraint：
\[
\left.t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n}\right\} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the decision boundary is allowed to be less than 1.
\[
1
\] decision boundary is alowedto be less than
\[
5-0.8+0
\]It says that the distance between a data point and the
\[
\operatorname{lec}_{2}+2
\] \(\square\)


\[
\begin{gathered}
\mathrm{w}^{\mathrm{T}} \mathrm{X}+\mathrm{w}_{0}=1 \\
\mathrm{w}^{\mathrm{T}} \mathrm{X}+w_{0}=0
\end{gathered}
\]




（
\(\square\)
portion margin sum －

號
\(\qquad\)電Remember the constraint below？
\(t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n\) ．

\section*{53̉ption 1：soft margin SVM}
\(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
\(\square\) For the data points which are non－separable，we relax
-
\[
\begin{gathered}
\mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=1 \\
\mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=0 \\
\mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=-1
\end{gathered}
\] the constraint：
\[
\begin{aligned}
& \quad t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{X}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0 \\
& \text { It says that the distance between a data point and the } \\
& \text { decision boundary is allowed to be less than } 1 \text {. }
\end{aligned}
\]
\[
\varepsilon_{n} \text { is called slack variables. }
\]
－\(\varepsilon_{n}\) is called slack variables．

（：sott margin SVM路 5

\section*{50ption 1：soft margin SVM}
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]

\section*{\((2)\)}
\(\qquad\)
```

                    O
    ```
＿
\(\qquad\)
．
\(\qquad\)
\(\square\)
\(\square\)

Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n}^{3} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
decision boundary is allowed to be less than 1.
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？ \(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
\(\square\) For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\) It says that the distance between a data point and the
decision boundary is allowed to be less than 1.
\(\square \varepsilon_{n}\) is called slack variables．
\(\square\) Question．Where is a data point when \(\varepsilon_{n}=1\) ？
\(\square\)
 \(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
\(\square\) For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n}: \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\) It says that the distance between a data point and the
decision boundary is allowed to be less than 1.
\(\varepsilon_{n}\) is called slack variables．
\(\square\) Question．Where is a data point when \(\varepsilon_{n}=1\) ？
\(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \text { ？} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
decision boundary is allowed to be less than 1.
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？ Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \text { ？} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
decision boundary is allowed to be less than 1 ．
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？ Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
decision boundary is allowed to be less than 1.
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？ \(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
\(\square\) For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\) It says that the distance between a data point and the
decision boundary is allowed to be less than 1.
\(\square \varepsilon_{n}\) is called slack variables．
\(\square\) Question．Where is a data point when \(\varepsilon_{n}=1\) ？
\(\square\)

\begin{abstract}

\end{abstract}

\(\square\) Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1: \varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
decision boundary is allowed to be less than 1 ．
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？




 \(\square\)



都 Remember the constraint below？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
decision boundary is allowed to be less than 1 ．
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1 ?\) Remember the constraint below？
\[
t_{n}\left(w^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n
\]
For the data points which are non－separable，we relax
the constraint：
\[
t_{n}\left(w^{\mathrm{T}} \mathrm{X}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
It says that the distance between a data point and the
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？
denary is allowed to be less than 1 ．
in

\[
\begin{aligned}
& \quad t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0 \\
& \text { It says that the distance between a data point and the } \\
& \text { decision boundary is allowed to be less than } 1 \text {. }
\end{aligned}
\]
decision boundary is allowed to be less than 1 ．
\(\varepsilon_{n}\) is called slack variables．
Question．Where is a data point when \(\varepsilon_{n}=1\) ？
.
.
 T
Remember the constraint below？
\(t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1, \quad \forall n\)
For the data points which are non－separable，we relax
the constraint：
\(t_{n}\left(\mathrm{w}^{\mathrm{T}} \mathrm{x}_{n}+w_{0}\right) \geq 1=\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0\)
者
```

* 

```
,
\(\qquad\)家

So we have the constraint below. How a
So we have the constraint below. How a
So we have the constraint below. How a




(soft margin SVM
soft margin SVM
\(\qquad\)
\(\qquad\)
soft margin SVM




Option \(1:\) soft margin SVM
(s) soft margin SVM



\[
5
\]
\(\Delta\)
\[
4^{k^{\prime}}
\]

\(\qquad\)

\[
\begin{aligned}
& \mathrm{w}^{\mathrm{T}} \mathbf{x}+w_{0}=1 \\
& \mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=0 \\
& \mathbf{w}^{\mathrm{T}} \mathbf{X}+w_{0}=-1
\end{aligned}
\]

\begin{abstract}
.

\end{abstract}
.


\section*{55ption 1: soft margin SVM \\ ( \\ }   


\section*{}

(s) soft ion
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} x_{n}+w_{0}\right) \geq-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\)
\(\square\) objective function?
\(\triangle\) objective function?
\(\qquad\)

\section*{Option 1：soft margin SVM \\ Option 1：soft marg SVI}
\(\square\) So we have the constraint below．How about the
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} x_{n}+w_{0}\right) \geq 1 \varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\) We want to minimize the slack．
\[
\begin{gathered}
\mathrm{w}^{\mathrm{T}} \mathrm{X}+w_{0}=1 \\
\mathrm{w}^{\mathrm{T}} \mathrm{X}+w_{0}=0 \\
\mathrm{w}^{\mathrm{T}} \mathrm{X}+w_{0}=-1
\end{gathered}
\]
亿

 \(+\)


\(\square\)

objective function？
\[
\min \frac{1}{2}\|\mathrm{w}\|^{2}+C \sum_{n} \varepsilon_{n}
\]
obj

.
.
正

\section*{\(+\)}

\[
0
\] ，



■
.
.
.
-

\footnotetext{

} （
（ \(\square\)共
－If＂ C ＂is small，the dominant factor is \(\|w\|^{2} / 2\)
  

So we have the constraint below．How about the objective function？
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} x_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\) We want to minimize the slack．
\(\min \frac{1}{2}\|\mathrm{w}\|^{2}+C \sum_{n} \varepsilon_{n}\)
.
－We want
objective function？正 ．
左
\(\qquad\)
\(\qquad\)
.





\(\square\)
.
\[
5
\]

1）Prefer large margin
2）May cause large \＃of misclassified data points．
路

\(+2\)


\section*{Option 1：soft margin SVM}

57


\section*{Option 1: soft margin SVM}


So we have the constraint below. How about the objective function?
\[
t_{n}\left(\mathrm{w}^{\mathrm{T}} x_{n}+w_{0}\right) \geq 1-\varepsilon_{n} \quad \forall n \quad \varepsilon_{n} \geq 0
\]
\(\square\) We want to minimize the slack.
\(\min \frac{1}{2}\|\mathrm{w}\|^{2}+C \sum_{n} \varepsilon_{n}\)
\(\square\) If " \(C\) " is small, the slack contributes more
1) Prefer large margin
2) May cause large \# of misclassified data points.
\(\square\) If "C" is large, the slack contributes less
1) Prefer less \# of misclassified data points.
2) May cause small margin.

Kernel trick
- They are the same problem.
- \(\lambda\) : Lagrange multipliers which corresponding to data points.
- t: label (-1 or 1)
- It looks complicated why we border to use dual problem???
\[
\begin{gathered}
\min _{\lambda} L(\lambda)=\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} t_{n} t_{m} \lambda_{n} \lambda_{m} \mathrm{x}_{n}^{T} \mathrm{x}_{m}-\sum_{n=1}^{N} \lambda_{n} \\
\text { s.t. } \lambda \geq 0, \quad t^{T} \lambda=0
\end{gathered}
\]

Dual problem
ipliers which corresponding
ted why we border to use
相
dual problem???
dual proble

Primal problem
Primal problem
\[
\begin{array}{ll}
\min & \frac{1}{2} \mathrm{w}^{T} \mathrm{w} \\
\text { s.t. } & t_{n}\left(\mathrm{w}^{\mathrm{T}} x_{n}+w_{0}\right) \geq 1
\end{array}
\]


 4
-



\section*{kernel trick ．}


 ．
\(\rightarrow \cos\)

 －
t linearly separable，what should we do？
\(\qquad\)




 ．



\section*{erne trick}


> f號

St．\(\lambda \geq 0, \quad t^{T} \lambda=0\)
\(\square\) If data \(\mathrm{x}_{\mathrm{n}}\) are not linearly separable，what should we do？
63
．號






St．\(\lambda \geq 0, \quad t^{T} \lambda=0\)
\(\square\) If data \(\mathrm{x}_{\mathrm{n}}\) are not linearly separable，what should we do？
63

St．\(\lambda \geq 0, \quad t^{T} \lambda=0\)
\(\square\) If data \(\mathrm{x}_{\mathrm{n}}\) are not linearly separable，what should we do？
63


St．\(\lambda \geq 0, \quad t^{T} \lambda=0\)
\(\square\) If data \(\mathrm{x}_{\mathrm{n}}\) are not linearly separable，what should we do？
63



 T
－
\(\min _{\lambda} L(\lambda)=\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} t_{n} t_{m} \lambda_{n} \lambda_{m} \mathrm{x}_{n}^{T} \mathrm{X}_{m}-\sum_{n=1}^{N} \lambda_{n}\)
Sit．\(\quad \lambda \geq 0, \quad t^{T} \lambda=0\)







\[
5-2-2
\]

\begin{abstract}
\(\min _{\lambda} L(\lambda)=\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} t_{n} t_{m} \lambda_{n} \lambda_{m} x_{n}^{T} x_{m}-\sum_{n=1}^{N} \lambda_{n}\)
sit．\(\lambda \geq 0, \quad t^{T} \lambda=0\)
\[
\begin{gathered}
\lambda \\
\vdots
\end{gathered}
\]

.
\end{abstract}
\(\qquad\)


\(\square\) If data \(x_{n}\) are not lin

\[
\begin{gathered}
\min _{\lambda} L(\lambda)=\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} t_{n} t_{m} \lambda_{n} \lambda_{m} z_{n}^{T} z_{m}-\sum_{n=1}^{N} \lambda_{n} \\
\text { sst. } \lambda \geq 0, \quad t^{T} \lambda=0
\end{gathered}
\]

\section*{正}
\(\qquad\)
```

* 

```
```

* 

```



\[
2
\]
\[
z_{1}
\]

\(\qquad\)

Kernel trick
\(\square\) The idea of Kernel trick begins from here：to find the scalar values（the inner product of two vectors：
\(z_{n}\) and \(z_{m}\) ）and so we can formulate the quadratic problem which can be linearly separable．
\(z_{n}\) and \(z_{m}\) ）and so we can formulate the quadratic problem which can be linearly separable． nab le 
        Kernel trick begins from here: to find the scalar values (the inner product of two vectors: Kernel trick begins from here：to find the scalar values（the inner product of two vectors：
and so we can formulate the quadratic problem which can be linearly separable．
.
:-....
\[
\begin{array}{r}
\text { Sf...................................................................... }
\end{array}
\]
\[
\square
\]

\(\square\) Kernel function \(K()\) is a function which returns the scalar values (the inner product of two vectors:
\(z_{n}\) and \(z_{m}\) in \(Z\) space) when the data points ( \(x_{n}\) and \(x_{m}\) in \(X\) space) are given.
\[
K\left(\mathrm{X}_{n}^{T}, \mathrm{X}_{m}\right)=\phi\left(\mathrm{X}_{n}^{T}\right) \phi\left(\mathrm{X}_{m}\right)=\mathrm{Z}_{n}^{T} \mathrm{Z}_{m}
\] \(K\left(\mathrm{X}_{n}^{T}, \mathrm{X}_{m}\right)=\phi\left(\mathrm{X}_{n}^{T}\right) \phi\left(\mathrm{X}_{m}\right)=\mathrm{Z}_{n}^{T} \mathrm{Z}_{m}\)
\[
\begin{aligned}
& \mathrm{z}_{n}^{T}=\phi\left(\mathrm{X}_{n}^{T}\right) \\
& \mathrm{z}_{m}=\phi\left(\mathrm{X}_{m}\right)
\end{aligned}
\]

\(\square\)

 
. ( \(+2\)
\[
x^{2}
\]
\[
\square
\]






\(\qquad\)
. －



Z spac （2）
 \(-\quad-2\)

\author{

}
\(\qquad\)
\(\qquad\)
\(\square\)







\footnotetext{
Z space problem can be formulated with data in X space
\(\square\)－

}

\section*{Z space problem \\ \(X\) space problem \\ espac priantor
 \\  \\ \[
\begin{gathered}
\min _{\lambda} L(\lambda)=\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} t_{n} t_{m} \lambda_{n} \lambda_{m} Z_{n}^{T} z_{m}-\sum_{n=1}^{N} \lambda_{n} \quad \square \quad \min _{\lambda} L(\lambda)=\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} t_{n} t_{m} \lambda_{n} \lambda_{m} \mathrm{~K}\left(\mathrm{x}_{n}^{T} \mathrm{x}_{m}\right)-\sum_{n=1}^{N} \lambda_{n} \\
\text { s.t. } \lambda \geq 0, \quad t^{T} \lambda=0 \\
\text { s.t. } \lambda \geq 0, \quad t^{T} \lambda=0
\end{gathered}
\] \\  \\ \(\square\) With the Kernel function defined previously，we want to change the quadratic problem as follows：
\(\quad-\quad\) Because the Kernel function is a function of data points（ \(x_{n}\) and \(x_{m}\) ）which we already have． \\  \\ \(\qquad\) \\ \\ \section*{```

```}} \\ \\ \section*{```

```}}


－
.
.
－ \(+\)


Tolynomial kernel of degree 2



\(\square\)
\[
\begin{aligned}
& \underbrace{\substack{\text { Space } \mathrm{x} \\
\left(y_{1}, y_{2}\right)}}_{\left(x_{1}, x_{2}\right)} \\
& K(\mathrm{x}, y)=(\mathrm{xy})^{2}
\end{aligned}
\]



\begin{abstract}
\(\qquad\)
\end{abstract}



_




\(\square\)
\(\square\)
正



F 2
\(\qquad\)
F
\[
\dot{\square}
\]
\[
K(\mathrm{x}, \mathrm{y})=(\mathrm{xy})^{2}
\] ，
er － 5
\(\qquad\)
\(\qquad\)
\(\qquad\)．－
\(\square\)
\(\qquad\)
\(\qquad\)
\(\square\)
\(\qquad\)

\footnotetext{
（ 1020
}

\begin{abstract}

\end{abstract}


\begin{abstract}
（
\end{abstract}

\footnotetext{
.
}
迤

．．

6̊olynomial kernel of degree 2

68
\(\qquad\)

\(\square\)
\(\square\)
\(\square\)

\begin{abstract}
\begin{abstract}
-
\end{abstract}
\end{abstract} . , ( \(\square\)

\begin{abstract}
-
\end{abstract}
 ( ( 
 

``` .
```

(2)

$$
\begin{aligned}
& \text { Space X } \\
& K(\mathrm{x}, \mathrm{y})=(\mathrm{xy})^{2} \\
& =\left(\left(x_{1}, x_{2}\right) \cdot\left(y_{1}, y_{2}\right)\right)^{2} \\
& =\left(x_{1} y_{1}+x_{2} y_{2}\right)^{2} \\
& =x_{1}^{2} y_{1}^{2}+2 x_{1} x_{2} y_{1} y_{2}+x_{2}^{2} y_{2}^{2} \\
& \xrightarrow[\substack{\text { Space } \mathrm{X}}]{\substack{\mathrm{x} \\
\left(x_{1}, x_{2}\right)}} \\
& \begin{array}{l}
\mathrm{y} \\
\Delta
\end{array} \\
& \text { }
\end{aligned}
$$

69olynomial kernel of degree 2


$$
\begin{align*}
K(\mathrm{x}, y) & =(\mathrm{xy})^{2} \\
& =\left(\left(x_{1}, x_{2}\right) \cdot\left(y_{1}, y_{2}\right)\right)^{2} \\
& =\left(x_{1} y_{1}+x_{2} y_{2}\right)^{2} \\
& =x_{1}^{2} y_{1}^{2}+2 x_{1} x_{2} y_{1} y_{2}+x_{2}^{2} y_{2}^{2}
\end{align*}
$$



[^0]```
                                *
```

```
                                *
```

Polynomial kernel of degree 2


Space X

$$
\begin{aligned}
K(\mathrm{x}, y) & =(\mathrm{xy})^{2} \\
& =\left(\left(x_{1}, x_{2}\right) \cdot\left(y_{1}, y_{2}\right)\right)^{2} \\
& =\left(x_{1} y_{1}+x_{2} y_{2}\right)^{2} \\
& =x_{1}^{2} y_{1}^{2}+2 x_{1} x_{2} y_{1} y_{2}+x_{2}^{2} y_{2}^{2}
\end{aligned}
$$



$$
\begin{aligned}
\phi(\mathrm{x}) \phi(y) & =\left(x_{1}^{2}, \sqrt{2} x_{1} x_{2}, x_{2}^{2}\right) \cdot\left(y_{1}^{2}, \sqrt{2} y_{1} y_{2}, y_{2}^{2}\right) \\
& =x_{1}^{2} y_{1}^{2}+2 x_{1} x_{2} y_{1} y_{2}+x_{2}^{2} y_{2}^{2}
\end{aligned}
$$

Mapping to 3-dimension

[^1]

## )

TGaussian Kernel: derivation (inner product in the infinite z space)

$$
K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right)=\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right)
$$

$K\left(\mathrm{x}_{\mathrm{n}}, \mathrm{x}_{\mathrm{m}}\right)=\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right)$

## $\qquad$路

教

$$
\begin{aligned}
K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right) & =\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}_{m}\right)
\end{aligned}
$$



教


## Gaussian Kernel: derivation (inner product in the infinite z space)


$\qquad$
-

$$
0^{2}+2
$$


$\qquad$

- $\begin{aligned} K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right) & =\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \quad \begin{array}{c}\text { Taylor series expansion of } \\ \text { an exponential function }\end{array} \\ & =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{X}_{m}\right) \quad \mathrm{x}+\frac{x}{0!}+\frac{x^{2}}{2!}+\frac{x^{3}}{3!}+\cdots \\ & =\exp \left(-\alpha \mathrm{X}_{n}^{2}\right) \exp \left(-\alpha \mathrm{X}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{X}_{\mathrm{n}}\right)^{k}\left(\mathrm{X}_{\mathrm{m}}\right)^{k}}{\mathrm{k!}}\end{aligned}$ $\begin{aligned} K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right) & =\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \quad \begin{array}{c}\text { Taylor series expansion of } \\ \text { an exponential function }\end{array} \\ & =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{X}_{m}\right) \quad \mathrm{x}+\frac{x}{0!}+\frac{x^{2}}{2!}+\frac{x^{3}}{3!}+\cdots \\ & =\exp \left(-\alpha \mathrm{X}_{n}^{2}\right) \exp \left(-\alpha \mathrm{X}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{X}_{\mathrm{n}}\right)^{k}\left(\mathrm{X}_{\mathrm{m}}\right)^{k}}{\mathrm{k!}}\end{aligned}$


## Gaussian Kernel: derivation (inner product in the infinite z space)

I

$$
\begin{aligned}
K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right) & =\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}_{m}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{x}_{\mathrm{n}}\right)^{k}\left(\mathrm{x}_{\mathrm{m}}\right)^{k}}{\mathrm{k}!}
\end{aligned}
$$




正
$\square$
.
.
. $K\left(\mathbf{X}_{n}, \mathbf{X}_{m}\right)=\exp \left(-\alpha\left\|\mathbf{X}_{n}-\mathbf{X}_{m}\right\|^{2}\right) \quad \begin{gathered}\text { Taylor series expansion of } \\ \text { an exponential function }\end{gathered} \quad \exp (x)=\frac{x^{0}}{0!}+\frac{x}{1!}+\frac{x^{2}}{2!}+\frac{x^{3}}{3!}+\cdots \quad 10$
$\square$
(




$\square$
-

$$
\begin{aligned}
K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right) & =\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}_{m}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{x}_{\mathrm{n}}\right)^{k}\left(\mathrm{x}_{\mathrm{m}}\right)^{k}}{\mathrm{k}!} \\
& =\sum_{k=0}^{\infty} \sqrt{\frac{(2 \alpha)^{k}}{\text { an exporsenenential expansion of of }}} \begin{array}{l}
\exp (x)=\frac{x^{0}}{0!} \\
\text { a! } \\
\\
\end{array} \frac{x}{1!}+\frac{x^{2}}{2!}+\frac{x^{3}}{3!}+\cdots
\end{aligned}
$$

## Gaussian Kernel: derivation (inner product in the infinite z space)

## Gaussian Kernel: derivation (inner product in the infinite z space)

$$
\begin{aligned}
& K\left(\mathrm{x}_{n}, \mathrm{x}_{\mathrm{m}}\right)=\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}_{m}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{x}_{\mathrm{n}}\right)^{k}\left(\mathrm{x}_{\mathrm{m}}\right)^{k}}{\mathrm{k}!} \\
& =\sum_{k=0}^{\infty} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{n}^{2}\right)\left(\mathrm{x}_{\mathrm{n}}\right)^{k} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{m}^{2}\right)\left(\mathrm{x}_{\mathrm{m}}\right)^{k} \\
& \text { Taylor series expansion of } \\
& \text { an exponential function } \\
& \exp (x)=\frac{x^{0}}{0!}+\frac{x}{1!}+\frac{x^{2}}{2!}+\frac{x^{3}}{3!}+\cdots \\
& \text { n exponential function } \\
& 1
\end{aligned}
$$

$\sum^{2}$

|  |
| :---: |
|  |
| $\left.x_{n}\right)^{k}$ |

$\square$
$\begin{aligned} K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right) & =\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}^{2}\right. \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}( }{} \\
& =\sum_{k=0}^{\infty} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{n}^{2}\right)\left(\mathrm{x}_{\mathrm{n}}\right)^{k} \sqrt{\frac{(2 \alpha)}{k}}\end{aligned}$

$$
\begin{aligned}
& \cdots \\
& +\cdots
\end{aligned}
$$

    \(\begin{aligned} & K\left(\mathrm{x}_{n}, \mathrm{x}_{m}\right)=\exp \left(-\alpha\left\|\mathrm{x}_{\mathrm{n}}-\mathrm{x}_{\mathrm{m}}\right\|^{2}\right) \\ &=\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}_{m}\right) \\ &=\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{x}_{\mathrm{n}}\right)^{k}\left(\mathrm{x}_{\mathrm{m}}\right)^{k}}{\mathrm{k!}} \\ &=\sum_{k=0}^{\infty} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{n}^{2}\right)\left(\mathrm{x}_{\mathrm{n}}\right)^{k} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{m}^{2}\right)\left(\mathrm{x}_{\mathrm{m}}\right)^{k} \\ & \text { anexpenenenenexatant function of } \\ & \exp (x)=\frac{x^{0}}{0!}+\frac{x}{1}+\frac{x^{2}}{2!}+\frac{x^{3}}{3!}+\cdots\end{aligned}\)
    



$\square$
$\qquad$

$$
\cdots
$$

$$
-+\cdots
$$

## TGaussian Kernel：derivation（inner product in the infinite z space）

$$
\begin{aligned}
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \exp \left(2 \alpha \mathrm{x}_{n} \mathrm{x}_{m}\right) \\
& =\exp \left(-\alpha \mathrm{x}_{n}^{2}\right) \exp \left(-\alpha \mathrm{x}_{m}^{2}\right) \sum_{k=0}^{\infty} \frac{(2 \alpha)^{k}\left(\mathrm{x}_{\mathrm{n}}\right)^{k}\left(\mathrm{x}_{\mathrm{m}}\right)^{k}}{\mathrm{k}!} \\
& =\sum_{k=0}^{\infty} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{n}^{2}\right)\left(\mathrm{x}_{\mathrm{n}}\right)^{k} \sqrt{\frac{(2 \alpha)^{k}}{k!}} \exp \left(-\alpha \mathrm{x}_{m}^{2}\right)\left(\mathrm{x}_{\mathrm{m}}\right)^{k} \\
& =\phi\left(\mathrm{X}_{\mathrm{n}}\right) \phi\left(\mathrm{X}_{\mathrm{m}}\right)
\end{aligned}
$$

$$
\begin{aligned}
& \text { app } \\
& )^{k}
\end{aligned}
$$

$$
\begin{aligned}
& \text { 五 } \\
& \text { 正 }
\end{aligned}
$$

$$
0
$$



$\square$

$\square$





$$
\alpha=1
$$




$$
\alpha=100
$$




# Hand-on Experience 

        Cola: Google Colaboratory
    $\square$ A web base free google cloud service
$\quad$ Jupyter Notebook with Google Drive
$\square$ You can even use GPU for free!
$\quad$ Good but it provides the best effort service
$\quad-\quad$ You must save your things in your google drive or somewhere
$\quad$ else.
$\square$
$\square$
$\quad$ Resource check
• !cat /proc/meminfo /proc/cpuinfo
• !dy -h
80
$\square$

$\square$




## 81Colab: Principal Component Analysis (PCA)



## Data loading: MNIST

$\square$ MNIST data set

- http://yann.lecun.com/exdb/mnist/
- Training data
- One single file (45M) which includes 60,000 hand digit images for training,
- One single file (59K) which includes corresponding labels.
- Testing data
- One single file (7.5M) which includes 10,000 hand digit images for testing,
- One single file ( 9.8 K ) which includes corresponding labels.



## Making your gdrive as a working directory

$\square$ Defining a root directory where you mount your gdrive
from google.colab import drive drive.mount('/gdrive/')

- /gdrive/My Drive/Colab Notebooks/
$\square$ Running time measurement
drivemoun(/gdive/)

```
import datetime
```

import datetime
before = datetime.datetime.now().timestamp()
before = datetime.datetime.now().timestamp()
after = datetime.datetime.now().timestamp()
after = datetime.datetime.now().timestamp()
print( "Time taken:", after - before)
print( "Time taken:", after - before)
...

```
...
```


## 88 Colab: Support Vector Machine (SVM) <br> 

1) Go to the Colab

- https://colab.research.google.com

2) Select "GITHUB" and copy the link below into

- https://github.com/suyongeum/PMLWS2018 WS2.git

3) Select the notebook in the list

- OCT_23_2018_SVM.ipynb-

4) Go to "Runtime" - "Change runtime type"

- Python 3
- GPU

5) Save it into your gdrive

- "File" - "Save a copy in Drive ..."

ository: ©

$$
\text { eum/PMLWS2018_WS2 } \quad \underline{\text { master } v}
$$


 1)





$\square$
$\qquad$
$\qquad$
$\qquad$
$\qquad$
$\qquad$

## 8Eolab: Feature extraction

1) Go to the Colab

- https://colab.research.google.com

2) Select "GITHUB" and copy the link below into

- https://github.com/suyongeum/PMLWS2018 WS2.git

3) Select the notebook in the list

- OCT_23_2018_Feature_extraction.ipynb

4) Go to "Runtime" - "Change runtime type"

- Python 3
- GPU

5) Save it into your gdrive

- "File" - "Save a copy in Drive ..."
(10 suve cory ordive...




digits $=$ load didit
data, label $=$ digit
drint
 (1797, 64

19. 
20. 

15
4


Ctul F10 function provided by numpy module
Ctritm
ST data: each digit hass 8 by 8 dim
data into two parts: data - its. shape of data
and its label and its label
and its label
 Restart and run all. Change runtime type ang Manage sessions

Pastory: $\mathbb{Z}$
Path


() ${ }^{\text {OCT_23_2018_PCA.ipyb }}$
() 0 CT_23_2018_SVMM.ipybb

- $\square$

Path
vCEL
.
-
-
-
-
5. 5. ๑. Ө.
5. 5. ๑. Ө.
5. 5. ๑. Ө.
5. 5. ๑. Ө.



 .



GPU
(

```
                        4, (a)
```


-



## Backup Slides

## ${ }^{87}$ Finally finally．．．




[^0]:    

[^1]:    ,

