

## 1. Introduction

Wednesday, May 5, 2021 1:02 PM

feature extraction  $\rightarrow$  CoD  $\downarrow$   
 $\hookrightarrow$  PCA  $\rightarrow$  f $\downarrow$



100 Days of ML

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discrtizati...

Day 33 - Working-...

Day 34 - Working...

Day 35 - Complete...

Day 36 - Handling...

Day 37 - Handling...

Day38-Missing Indi...

Day39 - KNN Impu...

Day40 - Iterative I...

Day 41 - Outliers in...

Day 42 - Outlier De...

Day 43 - Outlier de...

Day44 - Outlier De...

Day 45 - Feature C...

Day 46 - Curse of...

Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

5. Step By Step Solution

6. Coding the Steps

7. Practical Example on MNIST Dat...

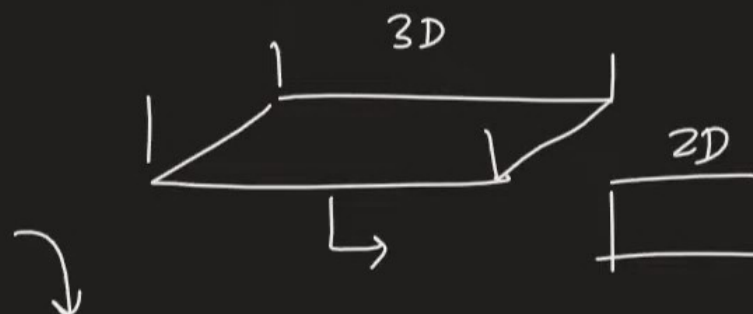
8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of Pr...

11. When PCA does not work

+ Add section + Add page



Benefits

1) faster exe of algo

2) Visualization

784

3D

10D

3D

2D



Wednesday, May 5, 2021 1:02 PM

No. of rooms	No. of grocery shops	Price (L)
3	2	60
4	0	130
5	6	170
2	10	90

$$d > d'$$

PCA

## feature selection

gymnastics

## Scatter

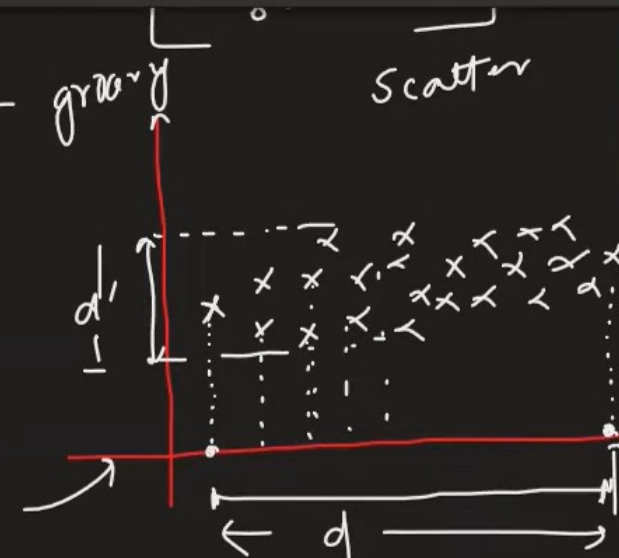
$X \rightarrow \text{rooms}$   
Variance

2008





rooms	grocery shops	(L)
3	2	60
4	0	130
5	6	170
2	10	90



variance

$$d > d'$$

No. of rooms | No. of washrooms | Price



$d = d'$   
feature extraction

## 100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

+ Add section

+ Add page



100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

5. Step By Step Solution

6. Coding the Steps

7. Practical Example on MNIST Dat...

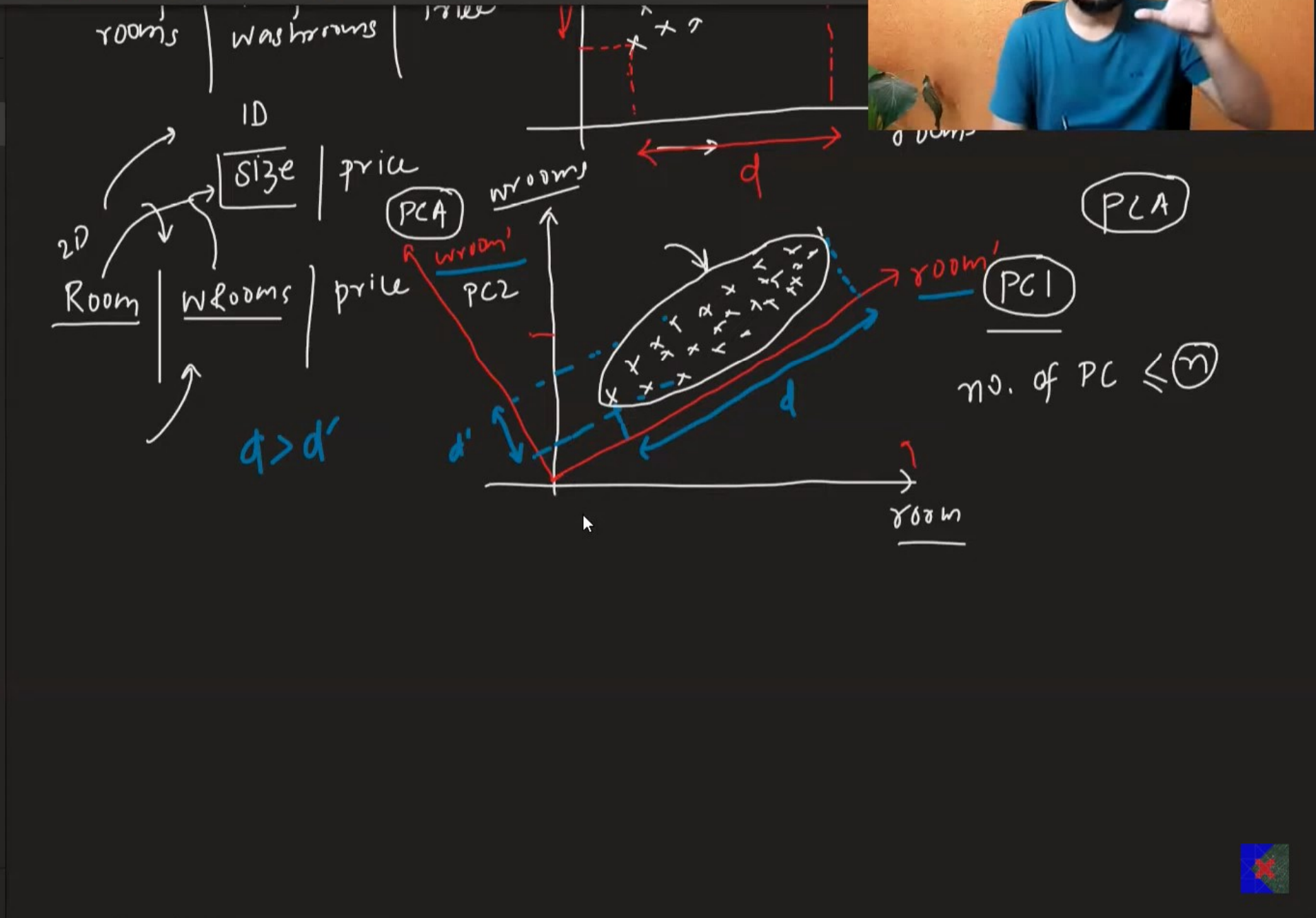
8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of Pr...

+ Add section

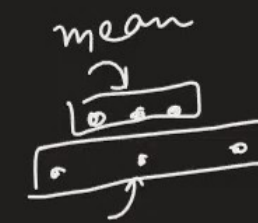
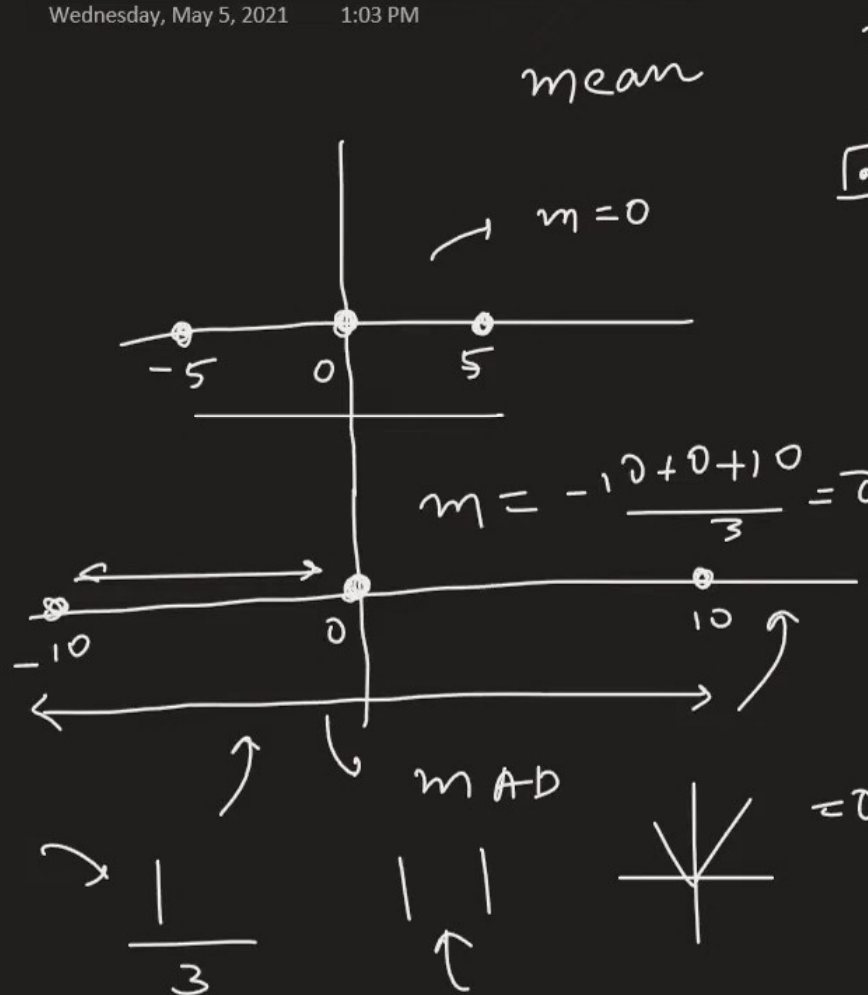
+ Add page





### 3. Why Variance is Important?

Wednesday, May 5, 2021 1:03 PM



variance

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

variance

$$v = \frac{25 + 0 + 25}{3} = \frac{50}{3}$$

$$v = \frac{100 + 0 + 100}{3} = \sqrt{\frac{200}{3}} = sd$$

#### 100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

+ Add section

+ Add page



←

→

OneNote for Windows 10

Home

Insert

Draw

View

Help

↶

↷

AI

+

↕

+

Shapes

Ink to Shape

Ink to Text

100 Days of ML

▼

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discrtizati...

Day 33 - Working-...

Day 34 - Working...

Day 35 - Complete...

Day 36 - Handling...

Day 37 - Handling...

Day38-Missing Indi...

Day39 - KNN Impu...

Day40 - Iterative I...

Day 41 - Outliers in...

Day 42 - Outlier De...

Day 43 - Outlier de...

Day44 - Outlier De...

Day 45 - Feature C...

Day 46 - Curse of...

Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

5. Step By Step Solution

6. Coding the Steps

7. Practical Example on MNIST Dat...

8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of Pr...

+ Add section

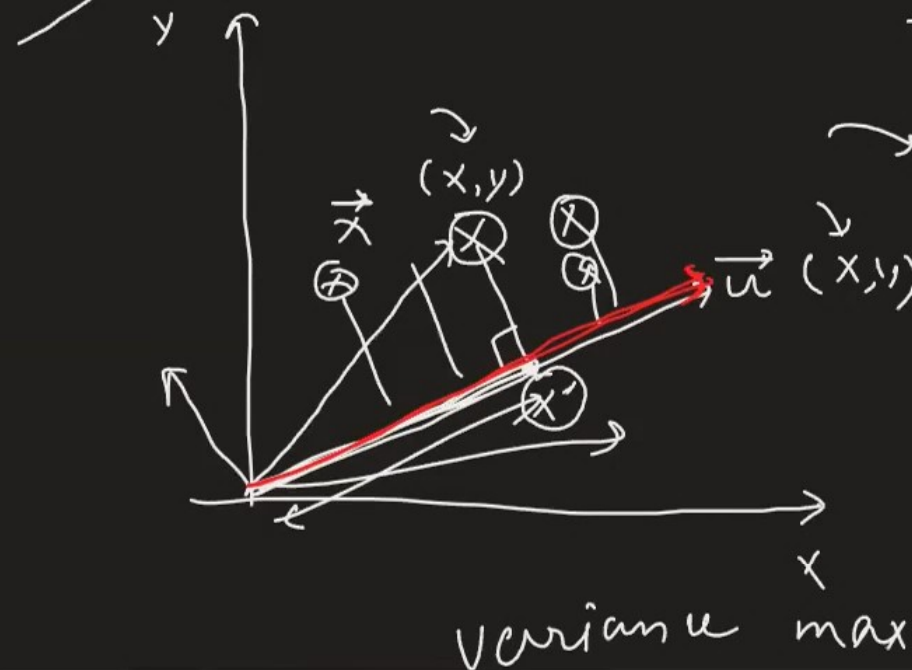
+ Add page

3. Why Variance is Important?

Wednesday, May 5, 2021 1:03 PM

The diagram illustrates Principal Component Analysis (PCA) on a 2D dataset. It shows a scatter plot of data points (marked with 'x') in a coordinate system with axes labeled 'x' and 'y'. A blue line represents the first principal component (PC1), which captures the maximum variance in the data. A red line represents the second principal component (PC2), which is orthogonal to PC1. The variance captured by each component is indicated by blue double-headed arrows along the respective axes. The word 'pca' is handwritten in the top right corner of the plot area.





$$\frac{\vec{u} \cdot \vec{x}}{|\vec{u}|=1} = \vec{u} \cdot \vec{x} = \boxed{u^T x}$$

$[x \ y] \ [x \ y]$   
 $x \quad y$   
 $[x_1 \ x_2] \ [y_1 \ y_2]$   
 $= [x_1 x_2 + y_1 y_2] - \text{scale}$





← →

Home Insert Draw View Help

↶ ↷ AI + -

Shapes Ink to Shape Ink to Text

100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction
2. Geometric Intuition
3. Why Variance is Important?
4. Problem Formulation
5. Step By Step Solution
6. Coding the Steps
7. Practical Example on MNIST Dat...
8. Visualization of MNIST Dataset
9. Explained Variance
10. Finding optimum number of Pr...
11. When PCA does not work

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

How to transform points?

variance maxi

unit vector

$\vec{u} \cdot \vec{x} = \vec{u} \cdot \vec{x} = \boxed{u^T x}$

$|\vec{u}|=1$

$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix}$

$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 & y_2 \end{bmatrix}$

$= \boxed{x_1 x_2 + y_1 y_2} - \text{scalar}$

OneNote for Windows 10

Home Insert Draw View Help

100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

5. Step By Step Solution

How to transform points?

6. Coding the Steps

7. Practical Example on MNIST Dat...

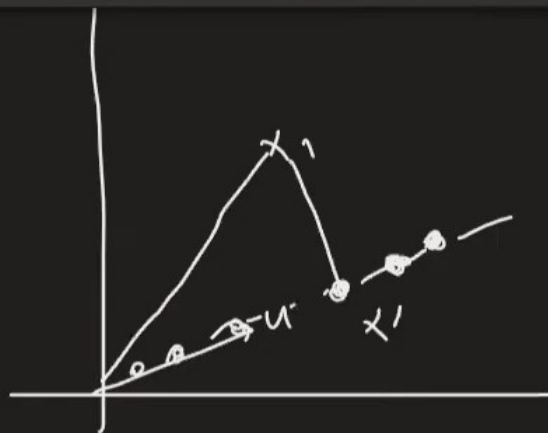
8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of Pr...

11. When PCA does not work

+ Add section + Add page



$$[u^T x_1]$$

$$\frac{1}{n} \sum_{i=1}^n (\overline{x_i} - \bar{x})^2$$

$$\frac{\sum_{i=1}^n (u^T x_i - u^T \bar{x})^2}{n}$$

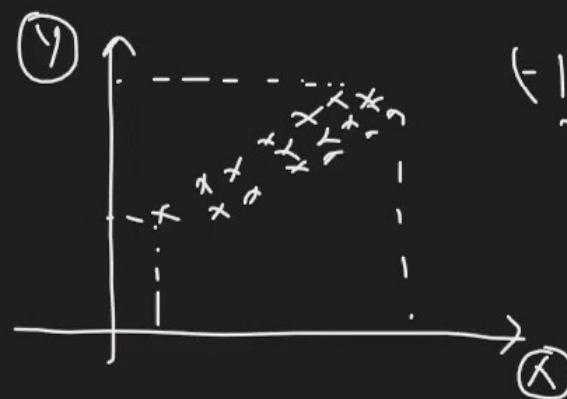
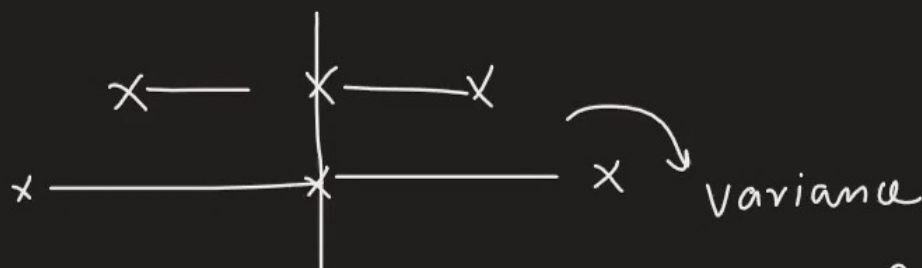
= variance  $\left( \frac{u}{\|u\|} \right)^2$





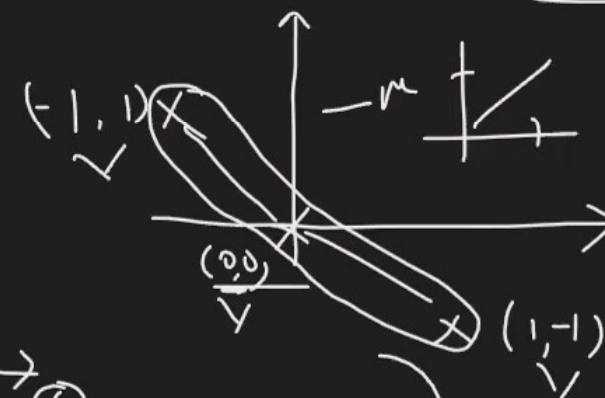
## Covariance and Covariance Matrix

Thursday, May 6, 2021 7:02 AM



Correlation

$-1$  to  $1$

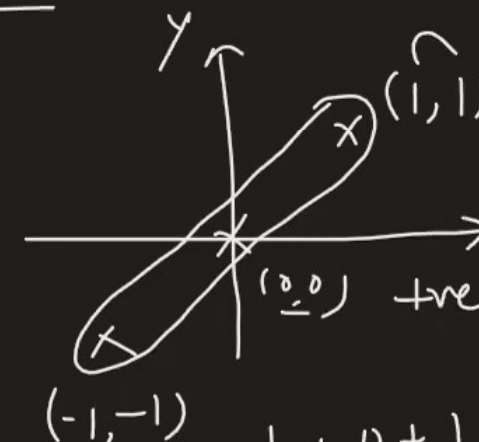


$$\frac{1 + 0 + 1}{3}$$

$$\left(-\frac{2}{3}\right)$$

$$\frac{-1 + 0 - 1}{3}$$

Covariance



$$\frac{1 + 0 + 1}{3}$$

$$= \left(\frac{2}{3}\right)$$





OneNote for Windows 10

Home Insert Draw View Help

100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

5. Step By Step Solution

How to transform points?

6. Coding the Steps

7. Practical Example on MNIST Dat...

8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of Pr...

11. When PCA does not work

+ Add section + Add page



$$\begin{array}{c|c} x_1 & x_2 \\ \hline x_1 & \\ x_2 & \end{array}$$

$$\text{cov mat}$$

$$\begin{bmatrix} 2 \times 2 \end{bmatrix}$$

$$\begin{array}{c} x_1 \\ x_2 \end{array} \begin{bmatrix} \text{cov}(x_1, x_1) & \text{cov}(x_2, x_1) \\ \text{cov}(x_1, x_2) & \text{cov}(x_2, x_2) \end{bmatrix}$$

$$\begin{array}{c} x_1 \\ x_2 \end{array}$$

$$\text{cov}(x_1, x_1) \rightarrow \text{var}(x_1)$$

$$\text{cov}(x_1, x_2)$$

$$\text{cov}(x_2, x_1)$$

$$\text{var}(x_2)$$

$$\rightarrow \text{square symmetric}$$

$$\text{cov}(a, b) = \text{cov}(b, a)$$

$$\begin{array}{c} v_x \\ v_y \\ v_z \end{array} \begin{array}{c} c_{xy} \rightarrow c_{xz} \\ c_{xy} \\ c_{xz} \end{array}$$

OneNote for Windows 10

Home Insert Draw View Help

100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

5. Step By Step Solution

How to transform points?

6. Coding the Steps

7. Practical Example on MNIST Dat...

8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of Pr...

11. When PCA does not work

+ Add section + Add page

$x \mid y \mid z$

$x$   $y$   $z$

$\sqrt{x}$   $C_{x,y}$   $C_{x,z}$

$C_{y,x}$   $\sqrt{y}$   $C_{y,z}$

$C_{z,x}$   $C_{z,y}$   $\sqrt{z}$

$x$   $y$   $z$

cov matrix





100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vect...

5. Step By Step Solution

How to transform points?

6. Coding the Steps

7. Practical Example on MNIST Dat...

8. Visualization of MNIST Dataset

9. Explained Variance

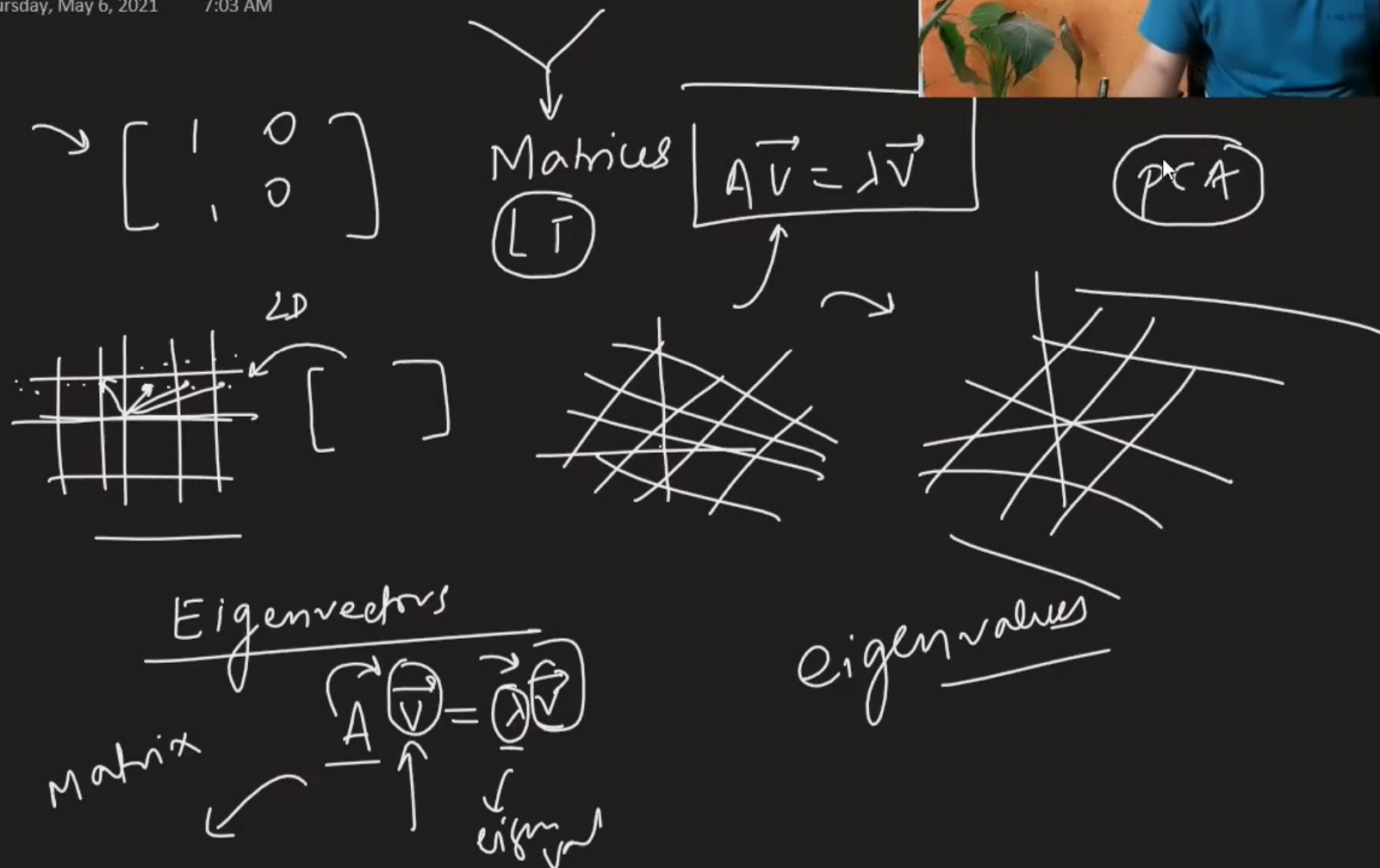
10. Finding optimum number of Pr...

11. When PCA does not work

+ Add section + Add page

## Linear Transformation, Eigen Vectors and Eigen Values

Thursday, May 6, 2021 7:03 AM



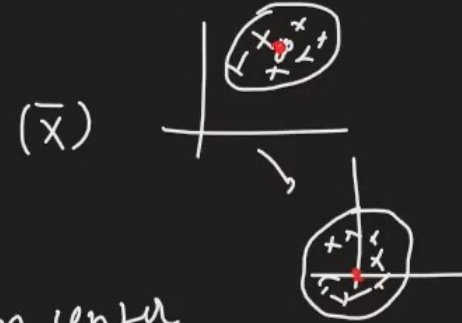




## 5. Step By Step Solution

Wednesday, May 5, 2021 1:03 PM

$f_1 \mid f_2 \mid f_3 \mid \dots$



1) mean center

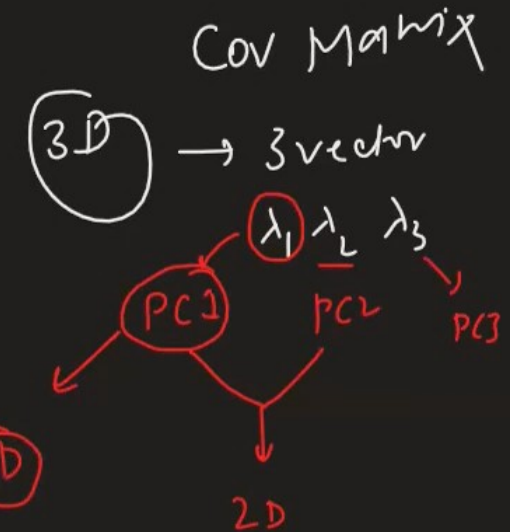
2) find cov matrix

3) find the eigen value/vector



$$\begin{matrix} f_1 \\ f_2 \\ f_3 \end{matrix} \begin{bmatrix} v(f_1) & c(f_1, f_2) & c(f_1, f_3) \\ c(f_1, f_2) & v(f_2) & c(f_2, f_3) \\ c(f_1, f_3) & c(f_2, f_3) & v(f_3) \end{bmatrix} \begin{matrix} f_1 \\ f_2 \\ f_3 \end{matrix}$$

cov mat  
(n p. cov)



### 100 Days of ML

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discrtizati...

Day 33 - Working-...

Day 34 - Working...

Day 35 - Complete...

Day 36 - Handling...

Day 37 - Handling...

Day38-Missing Indi...

Day39 - KNN Impu...

Day40 - Iterative I...

Day 41 - Outliers in...

Day 42 - Outlier De...

Day 43 - Outlier de...

Day44 - Outlier De...

Day 45 - Feature C...

Day 46 - Curse of...

Day 47 - PCA

+ Add section

+ Add page



100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

5. Step By Step Solution

How to transform points?

6. Coding the Steps

7. Practical Example on MNIST Dat...

8. Visualization of MNIST Dataset

9. Explained Variance

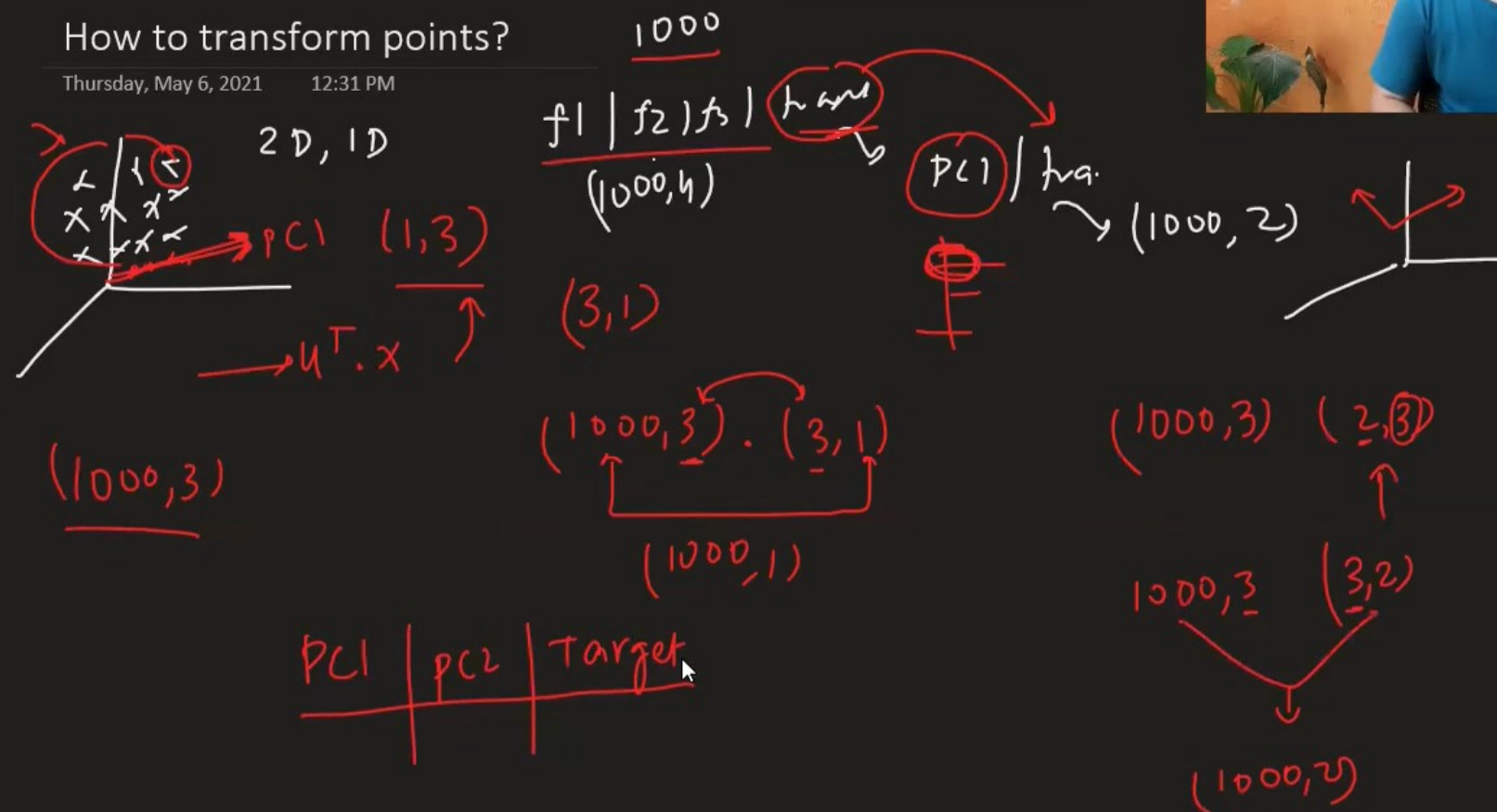
10. Finding optimum number of Pr...

11. When PCA does not work

+ Add section + Add page

## How to transform points?

Thursday, May 6, 2021 12:31 PM





## 10. Finding optimum number of Principle Components

Wednesday, May 5, 2021 1:05 PM

$EV = \lambda$

$(784) \rightarrow \lambda_1, \lambda_2, \lambda_3, \dots, \lambda_{784}$

$\left( \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_{784}} \right) \times 100 \rightarrow \text{percentage}$

$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \dots$

$\underline{30}, \underline{25}, \underline{15}, \underline{10}, \underline{5}$

$\lambda_{15}$

$\boxed{90\%}$

$\rightarrow \boxed{15} \lambda_1 - \lambda_{15}$

$90\%$

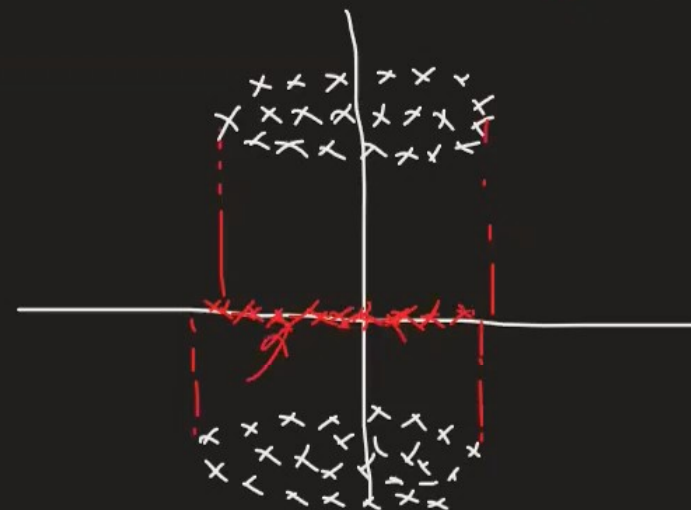
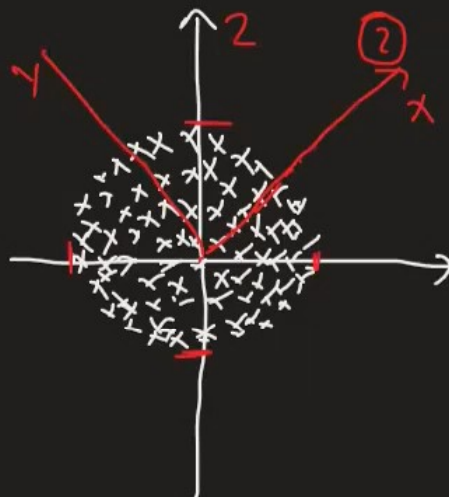




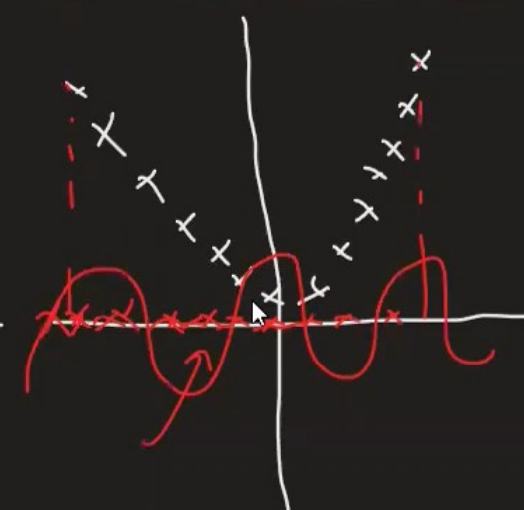


## 11. When PCA does not work

Thursday, May 6, 2021 7:03 AM



$$y = x^2$$



100 Days of ML

- Day 29 - Pipelines
- Day 30 - Function T...
- Day 31 - Power Tra...
- Day 32 - Discrtizati...
- Day 33 - Working-...
- Day 34 - Working...
- Day 35 - Complete...
- Day 36 - Handling...
- Day 37 - Handling...
- Day38-Missing Indi...
- Day39 - KNN Impu...
- Day40 - Iterative I...
- Day 41 - Outliers in...
- Day 42 - Outlier De...
- Day 43 - Outlier de...
- Day44 - Outlier De...
- Day 45 - Feature C...
- Day 46 - Curse of...
- Day 47 - PCA

1. Introduction

2. Geometric Intuition

3. Why Variance is Important?

4. Problem Formulation

Covariance and Covariance Matrix

Linear Transformation, Eigen Vecto...

5. Step By Step Solution

How to transform points?

6. Coding the Steps

7. Practical Example on MNIST Dat...

8. Visualization of MNIST Dataset

9. Explained Variance

10. Finding optimum number of P...

11. When PCA does not work

+ Add section + Add page