PCA & SVM ?

Prefix func^n ?

07 FRIDAY . FEBRUARY

NPTL
Pattern Recog.

Date: 06/08/17

FEBRUARY

Wk	M	T	W	T	F	S	S
5						1	2
6	3	4	5	6	7	8	9
7	10	11	12	13	14	15	16
8	17	18	19	20	21	22	23
9	24	25	26	27	28		

Data Mining

Data Preprocessing (Refer PPT).

What is data ?

* Collec^n of data objects and their attributes.

Typer of attributes

Text data

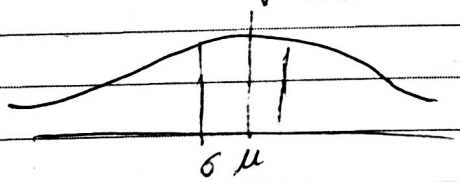
Tensor flow W2V

VC demension → Read.

Cogni'zance bias → book.

Central limit theorem → Sampling.

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



S_1 S_2 --- S_M

m_1 m_2 --- m_M

Attribute Transformation

A

0-1 normalization **

$\{a_1, a_2, \dots a_N\} \leftarrow [a_{min}, a_{max}] \leftarrow [a, b]$

$\{r_1, r_2 \dots r_n\} \leftarrow [0, 1]$

$$r_i = \frac{a_i - a_{min}}{a_{max} - a_{min}}$$

$$x_i = a + r_i(b-a)$$

2014

MARCH

Wk	M	T	W	T	F	S	S
--	31					1	2
10	3	4	5	6	7	8	9
11	10	11	12	13	14	15	16
12	17	18	19	20	21	22	23
13	24	25	26	27	28	29	30

SATURDAY · FEBRUARY | 08

8 |

9 |

10 |

11 |

12 |

1 |

2 |

3 |

4 |

5 |

Z transformation **

$x_1, x_2, \cdots x_N$; μ, σ

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

$$= \overline{x^2} - (\bar{x})^2$$

*

B_1	B_2
50	150
55	0
45	0
60	100
40	0

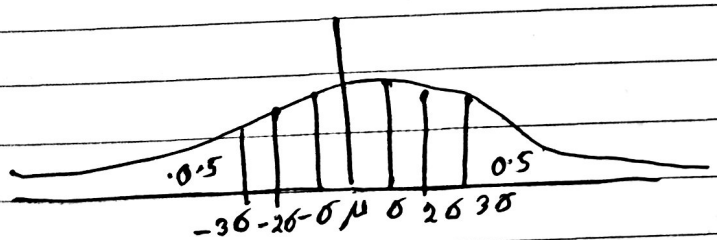
$$z_i = \frac{x_i - \mu}{\sigma}$$

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

SUNDAY | 09



2014

Imp → 0-1 & z trans.
→ distance

10 MONDAY · FEBRUARY

FEBRUARY

Wk	M	T	W	T	F	S	S
5						1	2
6	3	4	5	6	7	8	9
7	10	11	12	13	14	15	16
8	17	18	19	20	21	22	23
9	24	25	26	27	28		

· Similarety / Dessimilarity for sample Thru...

* Distance → Euclediαn distance.

$$d(\hat{x}_i, \hat{x}_j) = \left[\sum_{k=1}^{n} (x_{ik} - x_{jk})^2 \right]^{1/2}$$

* generalized dist.

$$d_r(\hat{x}_i, \hat{x}_j) = \left[\sum_{k=1}^{n} (x_{ik} - x_{jk})^r \right]^{1/r} \quad r > 0$$

$$d(\hat{x}_i, \hat{x}_j) = \sum_{k=1}^{n} |x_{ik} - y_{jk}|$$

Hamming dist.

① $d(\hat{x}, \hat{y}) \geq 0$

② $d(\hat{x}, \hat{x}) = 0$ if $\hat{x} = \hat{y}$

③ $d(\hat{x}, \hat{y}) + d(\hat{y}, \hat{z}) \geq d(\hat{x}, \hat{z})$

$$dis(\hat{x}, \hat{y}) = d(\hat{x}, \hat{y})$$

$$s(\hat{x}, \hat{y}) = \frac{1}{d(\hat{x}, \hat{y}) + 1}$$

$$= n - d(\hat{x}, \hat{y})$$

Cosine Similarity.

$$\hat{x}_i \cdot \hat{x}_j = |\hat{x}_i| \cdot |\hat{x}_j| \cos\theta$$

$$|\hat{x}_i| = \sqrt{x_{i1}^2 + x_{i2}^2 + \dots x_{in}^2}$$

$$\cos\theta = \frac{\hat{x}_i \cdot \hat{x}_j}{\|\hat{x}_i\| \cdot \|\hat{x}_j\|}$$

$$= \frac{\sum_{k=1}^{n} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{n} x_{ik}^2} \sqrt{\sum_{k=1}^{n} x_{jk}^2}}$$

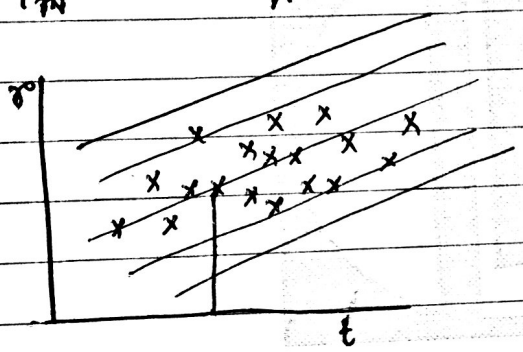
* Regression

x	y
Temp.	rainfall
t_1	r_1
t_2	r_2
:	
t_i	r_i
:	
t_N	r_N

$r = f(t)$

$r \propto t$

$r = ct$



$y = ax + b$

Act. value $r_i = at_i + b$

Pred. value R_i

$\sum_{i=1}^{N} |r_i - R_i|$

$SSE = \sum_{i=1}^{N} (r_i - R_i)^2$ → sqr. error

$MSE = \frac{1}{N} \sum_{i=1}^{N} (r_i - at_i - b)^2$

* Find a, b s.t. (so that) MSE is minimum ?

$f(x) = x^2 + 4x - 4$

$f'(x) = 0$ [maxima & minima]

$\Rightarrow 2x + 4 = 0$

$x = -2$

$f''(x) > 0$

$f(a,b)$

$$\frac{df}{da} = 0 \qquad \frac{df}{db} = 0$$

$$H(a,b) = \begin{vmatrix} \frac{d^2f}{da^2} & \frac{d^2f}{dadb} \\ \frac{d^2f}{dadb} & \frac{d^2f}{db^2} \end{vmatrix}$$

Hessian matrix

$\psi(a,b)$

$$\frac{\partial \psi}{\partial a} = \frac{1}{N}\sum_{i=1}^{N} 2(r_i - at_i - b)(-t_i) = 0$$

$$\frac{\partial \psi}{\partial b} = \frac{1}{N}\sum_{i=1}^{N} 2(r_i - at_i - b)(-1) = 0$$

$$\text{or, } \frac{1}{N}\sum_{i=1}^{N} r_i - a\cdot\frac{1}{N}\sum_{i=1}^{N} t_i - b = 0$$

$$\text{or, } \bar{r} - a\bar{t} - b = 0 \quad \text{——— (1)}$$

$$\overline{rt} - a\overline{t^2} - b\bar{t} = 0 \quad \text{——— (2)}$$

(2) − (1) × $\bar{t}$

$$\overline{rt} - \bar{r}\bar{t} - a(\overline{t^2} - \bar{t}^2) = 0$$

$$\boxed{a = \frac{\overline{rt} - \bar{r}\bar{t}}{\overline{t^2} - \bar{t}^2} = \frac{\mathrm{cov}(r,t)}{\mathrm{var}(t)}} = \frac{\mathrm{cov}(r,t)}{\sigma_t^2} = \frac{\mathrm{cov}(r,t)}{\sigma_r \sigma_t}\times\frac{\sigma_r}{\sigma_t}$$

$$a_{r,t} \quad \boxed{= \rho_{r,t}\frac{\sigma_r}{\sigma_t}}$$

$$\mathrm{Cov}(x,y) = \frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y}) = \overline{x_i y_i} - \bar{x}\cdot\bar{y}$$

$$\mathrm{Var}(x) = \frac{1}{N}\sum (x_i - \bar{x})^2$$

$$= \overline{x^2} - \bar{x}^2$$

→ Boltzman
→ cantor
dangerous knowledge youtube → Goedel
→ Touring

MARCH

Wk	M	T	W	T	F	S	S
—	31					1	2
10	3	4	5	6	7	8	9
11	10	11	12	13	14	15	16
12	17	18	19	20	21	22	23
13	24	25	26	27	28	29	30

SATURDAY . FEBRUARY 15

Corelation

$$Cor(x,y) = \rho(x,y)$$

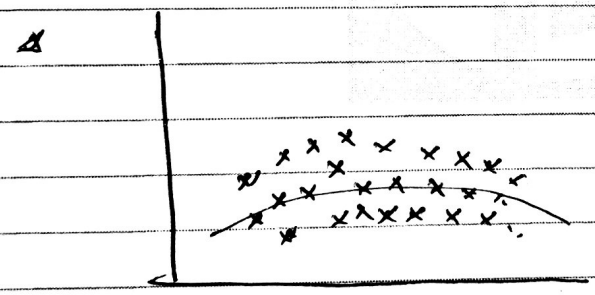
$$= \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

$$r = at + b \longrightarrow (3)$$

(3) − (1)

$$\boxed{r - \bar{r} - a(t - \bar{t}) = 0}$$

$$y - y_1 = m(x - x_1)$$



$$r = at^2 + bt + c$$

$$R_i = at_i^2 + bt_i + c$$

$$MSE = \frac{1}{N} \sum (r_i - at_i^2 - bt_i - c)^2$$ | Partial diff.ⁿ

SUNDAY 16

Pres	Temp	rainfall
P_1	t_1	r_1
P_2	t_2	r_2

$$r = a_0 + a_1 x_1 + a_2 x_2 \cdots + a_n x_n$$

2014