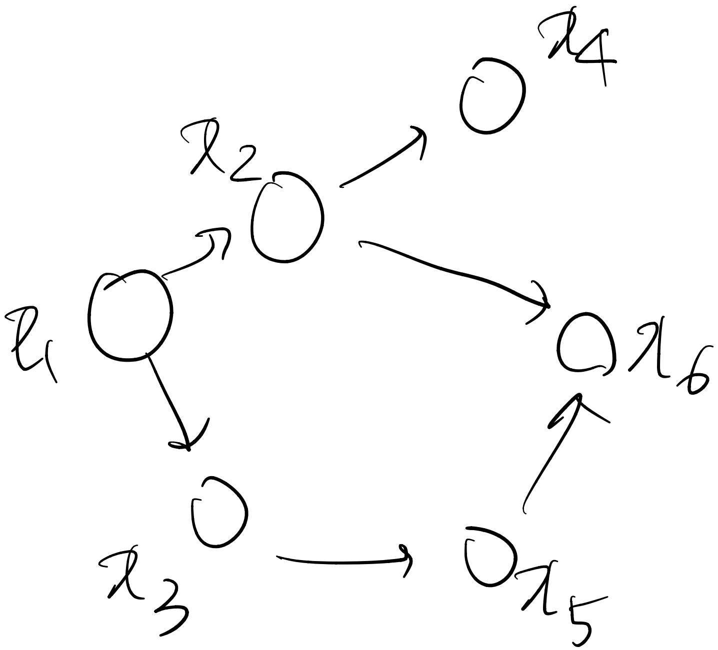
Directed Graphical Model.



$$p(X) = p(x_1, x_2, \cdots, x_6)$$

$$= p(x_1)\, p(x_2 | x_1)\, p(x_3 | x_1)$$

$$\cdot\, p(x_4 | x_2)\, p(x_5 | x_3)\, p(x_6 | x_2, x_5)$$

$$p(x) = p(x_1)\, p(x_2|x_1)\, p(x_3|x_1)\, p(x_4|x_2)\, p(x_5|x_3)\, p(x_6|x_2, x_5)$$

① for $p(x)$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ & & \sigma_3^2 & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ & & & \sigma_4^2 & \sigma_{45} & \sigma_{46} \\ & & & & \sigma_5^2 & \sigma_{56} \\ & & & & & \sigma_6^2 \end{bmatrix}$$

$$6 + \frac{6 \cdot 7}{2} = 27$$

---

② $p(x_1) \sim N(\mu_1, \sigma_1^2)$

$p(x_2|x_1) \sim N(\mu_{2|1}, \sigma_{2|1}^2)$

$$\mu_{2|1} = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x_1 - \mu_1)$$

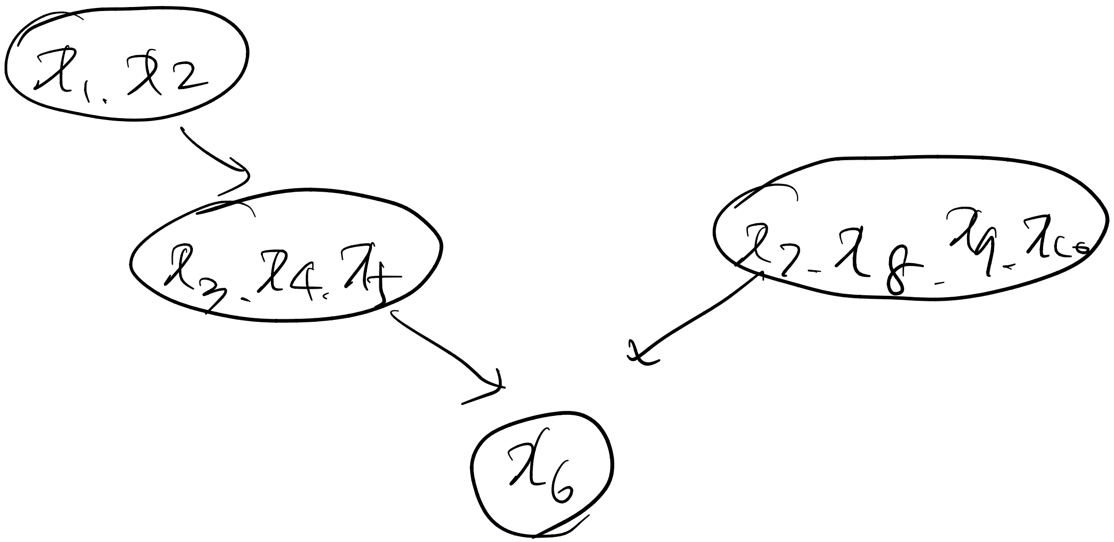
$$\sigma_{2|1}^2 = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}$$

$p(x_6|x_2, x_5) \sim N(\mu_{6|25}, \sigma_{6|25})$

$$\mu_{6|25} = \mu_6 + \Sigma_{6,25}\, \Sigma_{25}^{-1} \left(\begin{pmatrix} x_2 \\ x_5 \end{pmatrix} - \begin{pmatrix} \mu_2 \\ \mu_5 \end{pmatrix} \right)$$

$$(\sigma_{62}, \sigma_{65})$$

$$\begin{bmatrix} \sigma_5^2 & \sigma_{25} \\ \sigma_{52} & \sigma_5^2 \end{bmatrix}^{-1}$$



$$P(x_1, x_2, \cdots, x_{10}) = P(x_1, x_2)\, P(x_3, x_4, x_5 \mid x_1, x_2)\, P(x_7, x_8, x_9, x_{10})\, P(x_6 \mid x_3, x_4, x_5, x_7, x_8, x_9, x_{10})$$

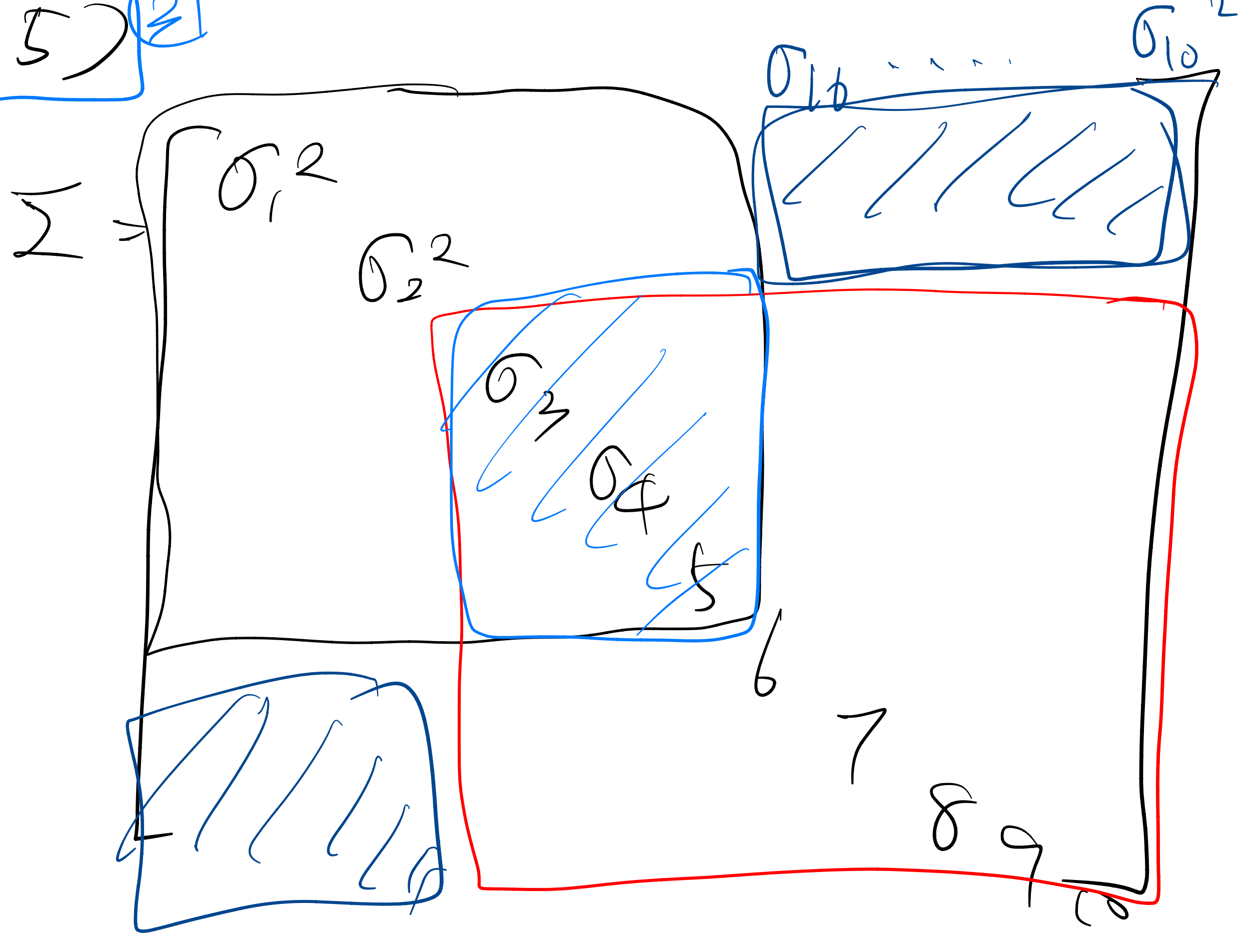
$$= \cancel{P(x_1, x_2)} \cdot \frac{P(1,2,3,4,5)}{\cancel{P(1,2)}} \cdot \cancel{P(7,8,9,10)} \cdot \frac{P(3 \sim 10)}{P(3,4,5,7,8,9,10)}$$

$$= P(3,4,5)\, \cancel{P(7,8,9,10)}$$

$$= \frac{P(1,2,3,4,5) \cdot P(3-10)}{P(3,4,5)}$$

① $\dfrac{P(1,2,3,4,5)\cdot P(3\text{—}10)}{P(3,4,5)}$ ②

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{10} \end{bmatrix}$$



Full Gaussian

$$10 + \frac{10\cdot 11}{2} = 65$$

Reduced

$$10 + \left(\frac{5\cdot 6}{2} + \frac{8\cdot 9}{2} - \frac{3\cdot 4}{2}\right) = 55$$

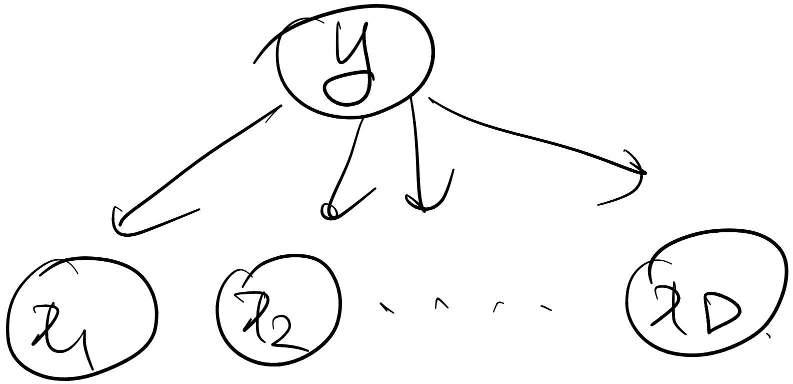
Naive Bayes

All random var. are mutually independent.

$$P(x) = \prod_{d=1}^{D} P_d(x_d) = P_1(x_1)P_2(x_2)\cdots P_D(x_D)$$

---

with class label y.



<Spam Classifier>

given an email T → Classify spam or not
$y=1$ $y=0$

① input definition (Features)

X_1 : Email contains the word "free"

X_2 : " "meeting"

X_3 " "Click"

⇒ binary discrete R.V

$$\underbrace{P(Y=\text{spam} \mid X_1, X_2, X_3)} = \frac{\boxed{P(X_1, X_2, X_3 \mid Y)}\, P(Y)}{P(X_1, X_2, X_3)}$$

X_1	X_2	X_3	$Y=\text{Spam}$	
0	0	0	1 →	1/60
0	0	1	1 →	1/60
0	1	0	1 →	1/10
0	1	1	0	1
1	0	0	3	1
1	0	1	1	
1	1	0	2	1
1	1	1	1	-

$$P(X_1, X_2, X_3 \mid Y) = P(X_1 \mid Y)\, P(X_2 \mid Y)\, P(X_3 \mid Y)$$

$P(X_1 \mid Y) \in$

X_1	$Y = \text{spam}$
1	
0	

X_1	$Y = \text{not spam}$
1	
0	

4

12

① construct a feature vector x of T.

Construct a dictionary

$$\left.\begin{bmatrix} \text{a} \\ \text{apple} \\ \vdots \\ \text{ze} \end{bmatrix}\right\} 10{,}000$$

binary feature vector

$x \in \{0,1\}^n$.

$n = 10{,}000$

$x_i = \mathbb{1}\{\text{word } i \text{ appears in } T\}$.

$$T \longrightarrow x = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

for $Y=1$ $\quad 1 \quad 2 \quad \cdots \quad 10.000$

$2^{10.000}$

$\boxed{2 \times 2^{10.000}}$

4

$$\Rightarrow P(x_1, x_2, \ldots, x_{10000} \mid y) = \boxed{P(x_1 \mid y)}\, P(x_2 \mid y) \cdots P(x_{10000} \mid y)$$

$\boxed{4 \times 10.000}$