SUMMARY OF PREV. LECT.

- Conditional Independence:

X is conditionally independent of Y given Z if

$$Prob(X \mid Y Z) = Prob(X \mid Z)$$

- Bayes Rule:

$$P(Y=y_k \mid X_1 \ldots X_n) = \frac{P(Y=y_k)\, P(X_1 \ldots X_n \mid Y=y_k)}{\sum_j P(Y=y_j)\, P(X_1 \ldots X_n \mid Y=y_j)}$$

[exponential est.]

Assuming Conditional Independence

$$P(Y=y_k \mid X_1 \ldots X_n) = \frac{P(Y=y_k) \prod_{i=1}^{n} P(X_i \mid Y=y_k)}{\sum_j P(Y=y_k) \prod_{i=1}^{n} P(X_i \mid Y=y_j)}$$

[linear est.]

- Naive Bayes Classifier:

↳ Training Rule: → MLE Estimates

$P(Y=y_k) = ?$, $P(X_i \mid Y=y_k) = ?$

↳ Classification Rule: $X^{new} = \langle X_1^{new}, X_2^{new}, \ldots, X_n^{new} \rangle$

$$Y_{new} = \underset{y_k}{\operatorname{argmax}}\; P(Y=y_k) \prod_{i=1}^{n} P(X_i^{new} \mid Y=y_k)$$

- Example Applications → Spam Filtering, Attachment Check → News Grouping, Text classify ....

1000 words / mail ⇐ max

Ex

$X_1 = \{$"I", "am", "Congrats" ....$\}$

$X_2 = \{$"You", "I", "am" ....$\}$

$X_3 =$

50000 words from Voc.

Hi,

Congrats! You won $1000000.

TE

S / N

MLE

Train:

50000

$P(x_i | S)$

$= P(x_j | S)$ ("i.i.d")

✓ $P($"$x_1$ = Congrats" $| Y = \text{Spam}) = ?$ ✓

✓ $P($"$x_1$ = Congrats" $| Y = NS) = ?$ ✓

$P(Y=S) = ?$ ✓

$P(Y=NS) = 1 - P(Y=S)$

New Mail

→ $x_1 = I$

→ $x_2 = \text{am}$

$x_3 = \text{pleased}$

$x_4 = \text{to}$

$x_5 = \text{inform}$

. . . .

$P($"$X_1 = I$" $| S) = ?$

$P($"$X_1 = I$" $| NS) = ?$

$P(x_2 = \text{am} | S) = ?$

$P(x_2 = \text{am} | NS) = ?$

✓

$P(Y = S | X_1 \ldots X_n) = ?$

$P(Y = NS | X_1 \ldots X_n) = ?$

high

S / NS

## Issues:

① Dependence on Conditional Independence.

$$y^{new} \leftarrow \underset{y_k}{\operatorname{argmax}} \; P(Y=y_k) \prod_{i=1}^{n} P(X_i^{new} \mid Y=y_k)$$

$$P(x_1 \mid Y=y_k) \cdot P(x_2 \mid Y=y_k) \cdots$$

$$\left\{ \begin{array}{l} x_1 = I \\ x_2 = am \end{array} \right. \qquad x_1 \neq \text{"You"}$$

---

② $P(x_i = a \mid Y = y_k)$ ← $\boxed{\text{MLE}}$

$$+ \quad = \frac{0}{3} = 0$$

| $x_i$ | |
|---|---|
| ~~a~~ | + |
| ~~b~~ | + |
| ~~d~~ | − |
| ~~c~~ | + |
| ~~e~~ | − |

TE

MAP

Soln.

Test; $(x_i = a) \longrightarrow$ Prob = 0, $(y+)$

$\langle x_1 \ldots x_i \ldots x_n \rangle$
b c. a . . . . $\longrightarrow$ $y = -ve$ ← may not be universally true.

$$\max \left\{ \begin{array}{ll} P(Y=+) \prod_{i=1}^{n} P(x_{i=a}^{new} \mid Y=+) = \checkmark & 0 \\ P(Y=-) \prod_{i=1}^{n} P(x_i^{new} \mid Y=-) = \checkmark & \neq 0 \end{array} \right.$$

MLE

$$\hat{\pi}_k = \hat{P}(Y = y_k) = \frac{\# \mathcal{D}\{Y = y_k\}}{|\mathcal{D}|}$$

$$\hat{\theta}_{ijk} = \hat{P}(X_i = x_{ij} \mid Y = y_k) = \frac{\# \mathcal{D}\{X_i = x_{ij} \wedge Y = y_k\}}{\# \mathcal{D}\{Y = y_k\}}$$

$\boxed{\text{Prior}} \rightarrow$ Dirichlet Distribution

MAP

$$\hat{\pi}_k = \hat{P}(Y = y_k) = \frac{\# \mathcal{D}\{Y = y_k\} + (\beta_k - 1)}{|\mathcal{D}| + \sum_m (\beta_m - 1)}$$

$$\hat{\theta}_{ijk} = \hat{p}(X_i = x_{ij} \mid Y = y_k) = \frac{\# \mathcal{D}\{x_{ij} \wedge y_k\} + (\beta_k - 1)}{\# \mathcal{D}\{y_k\} + \sum_m (\beta_m - 1)}$$

Test:

$$Y^{new} = \underset{y_k}{\operatorname{argmax}} \; \hat{\pi}_k \prod_{i=1}^{n} \hat{\theta}_{ijk}$$

Issues:

③ ~~Discrete~~ Continuous $X_i$ & Discrete $Y_k$ (classes) ✓

↳ ? (variation in Naive Bayes)

Solution: Gaussian Naive Bayes

$$P(X_i = x \mid Y = y_k) = \frac{1}{\sqrt{2\pi\sigma_{ik}^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_{ik}}{\sigma_{ik}}\right)^2}$$

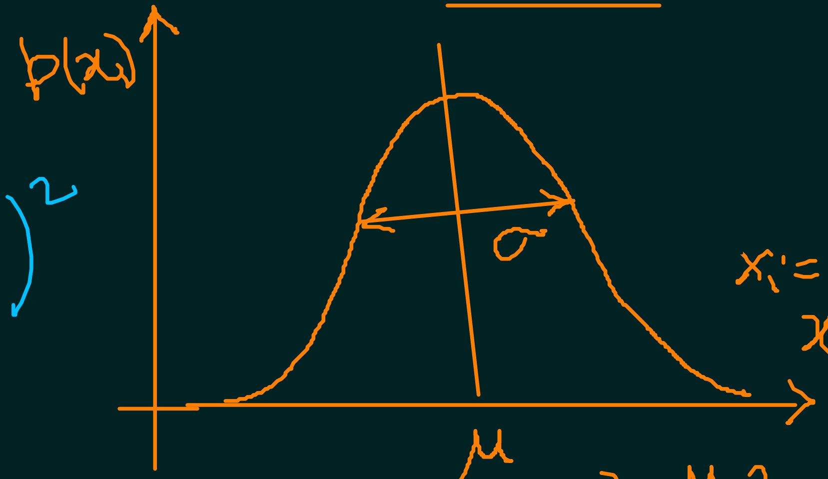
$$\left\{ \begin{array}{l} \text{conditional } \sigma_{ik} \\ \text{conditional } \mu_{ik} \end{array} \right\}$$

$\forall\ \sigma_{ik} = \sigma_k$ or $= \sigma_i$

Test: $Y^{new} \leftarrow \underset{y_k}{\operatorname{argmax}}\ P(Y = y_k) \prod_{i=1}^{n} \mathcal{N}(X_i^{new}, \sigma_{ik}, \mu_{ik})$

(Classify)

Gaussian

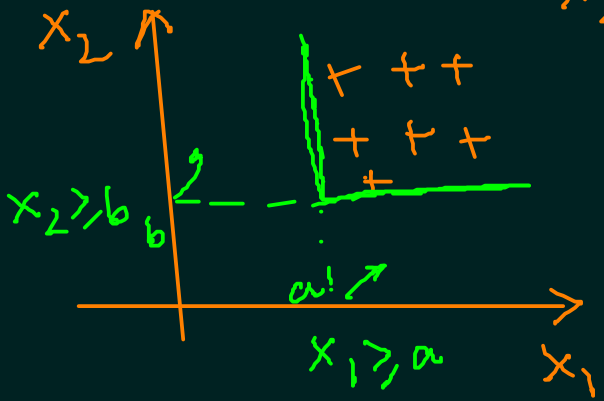


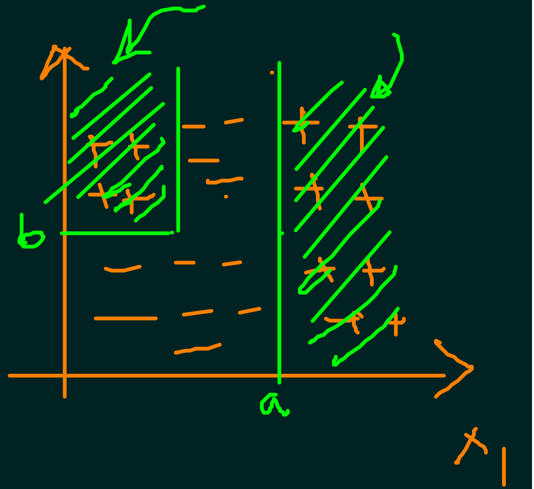

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\left\{ \begin{array}{l} \text{Mean} = \mu \\ \text{Var} = \sigma^2 \\ \text{SD} = \sigma. \end{array} \right\}$$

▷ **Ramification** : Naive Bayes

$$\frac{P(y=1 \mid X_1 \cdots X_n)}{P(y=0 \mid X_1 \cdots X_n)} \gtrless 1 \, ?$$




$$\Rightarrow \frac{P(y=1) \prod_{i=1}^{n} P(x_i \mid y=1)}{P(y=0) \prod_{i=1}^{n} P(x_i \mid y=0)} \gtrless 1$$

$(x_1 \geq a) \wedge (x_2 \geq a)$

$$\Rightarrow 0 \gtrless \underbrace{\ln\left[\frac{P(y=1)}{P(y=0)}\right]}_{\text{(constant)}} + \sum_{i=1}^{n} \ln\left[\frac{P(x_i \mid y=1)}{P(x_i \mid y=0)}\right]$$

Log linear classifier

HW

If $X_i = \{0,1\}$ then the threshold $(\gtrless 0)$ is a LINEAR Function of $X_i$'s.

$\Rightarrow$ Hint : $P(x_i = 0 \mid y=1) = 1 - P(x_i = 1 \mid y=1)$

# Gaussian Naive Bayes :

$P(\overset{(BB)}{\text{Sports}} \mid \text{Hight}, ML) = ?$

$P(H \mid S = \text{Good}) = GD \rightarrow P(H \mid S = \text{Bad}) = GD$

$P(ML \mid S = \text{Good}) = GD \rightarrow P(ML \mid S = \text{Bad}) = GD$

$P(S = \text{Good})\, P(H \mid S = \text{Good})\, P(ML \mid S = \text{Good})$