Probabilistic Classifiers

problem: $(x_1, \dots, x_n)$ data points / instances

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$$

$$c = \{c_1, \dots, c_k\} \qquad c_1 = 1, \dots, c_k = 10$$

$$D = \{(x_i, c_i) : i = 1, \dots, n\}$$

goal: given x^* find c^*

Prob. clf: $p(c_k \mid x_*) =$

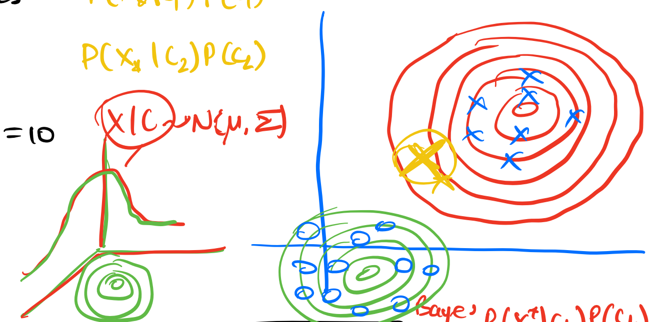
$$\boxed{c^* = \arg\max_k P(c_k \mid x_*)}$$

$$k=2: \quad c^* = \arg\max \{P(c_1 \mid x_*), P(c_2 \mid x_*)\}$$

$P(x_* \mid c_1) P(c_1)$

$P(x_* \mid c_2) P(c_2)$

$X \mid C \sim N(\mu, \Sigma)$



$$c^* = \boxed{\arg\max_k P(c_k \mid x^*)} \overset{\text{Bayes}}{=} \frac{P(x^* \mid c_k) P(c_k)}{P(x^*)}$$

How do we model $P(c_k \mid x^*)$

① Discriminative: directly model

$P(c_k \mid x^*)$ e.g. linear Regression

(x_i, y_i)

$y_i \mid x_i \sim N(x_i^T\beta, \sigma^2)$

② Generative: learn $P(x \mid c)$, $P(c)$

class conditional dist — class point

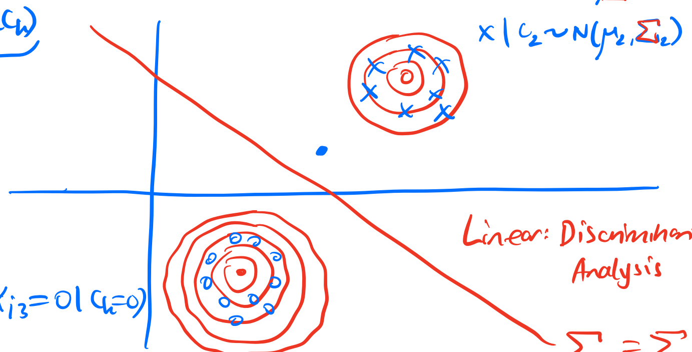
$$c_k = \arg\max_k \underbrace{P(c_k \mid x)}$$

$$= \arg\max_k \frac{\{P(x \mid c_k) P(c_k)\}}{\cancel{P(x)}} = \max \left\{ \frac{P(x \mid c_1) P(c_1)}{\cancel{P(x)}}, \frac{P(x \mid c_2) P(c_2)}{\cancel{P(x)}} \right\}$$

$$P(x) = \sum_{k=1}^{K} \underbrace{P(x \mid c_k) P(c_k)}$$

$x \mid c_1 \sim N(\mu_1, \Sigma_1)$

$x \mid c_2 \sim N(\mu_2, \Sigma_2)$



Linear Discriminant Analysis

$\Sigma_1 = \Sigma_2$

spherical

$P(x_{i3} \mid c_k) = ?$

$P(x_{i3} = 0 \mid c_k = 0)$ •

$P(x_{i3} = 1 \mid c_k = 0) = 1 - P(x_{i3} = 0 \mid c_k = 0)$

$P(x_{i3} = 0 \mid c_k = 1) \leftarrow$

$P(x_{i3} = 1 \mid c_k = 1)$

$P(x \mid c_k)$ continuous / discrete

normal — Multinomial, Multivariate Bernoulli

$p = 100, \ 2^{100}$

$$P(AB \mid c) P(c) = P(ABC)$$
$$= P(A \mid BC) P(BC)$$
$$= P(A \mid BC) \underbrace{P(B \mid C) P(C)}$$

$$P(X_i|C_k)P(C_k) = P(X_{i1}, X_{i2}, \dots X_{ip}|C_k)P(C_k)$$

$$= \underbrace{P(X_{i1}|X_{i2},\dots,X_{ip},C_k)}_{2^p} P(X_{i2}|X_{i3},\dots,X_{ip},C_k), \dots, P(X_{i,p-1}|X_{ip},C_k)$$

$$\times P(X_{ip}|C_k)P(C_k)$$

How to estimate? $X_{ij} \in \{0,1\}$ $C_k \in \{0,1\}$ $p=3$

$$P(X_i|C_k) = \underbrace{P(X_{i1}|X_{i2},X_{i3},C_k)}_{2^3} \underbrace{P(X_{i2}|X_{i3},C_k)}_{2^2} \underbrace{P(X_{i3}|C_k)}_{2} \underbrace{P(C_k)}_{1}$$

i	X_{i1}	X_{i2}	X_{i3}	C_i
1	1	0	0	1
2	0	0	0	0
3	1	1	1	0
4				

$P(C_k=1)$, $P(C_k=0)$

$$\frac{\#\text{rows with } c_i=1}{\text{total } \#\text{rows}}$$

$P(C_k=0) = 1-P(C_k=1)$

- $P(X_i|C_k)P(C_k) = P(X_{i1}|X_{i2},\dots,X_{ip},C_k), \dots, P(X_{ip}|C_k)P(C_k)$

Naive Bayes: conditional independence of features

X $x_{i1} \perp x_{i2}$ $j \neq m$ $\boxed{P(X_{i1},X_{i2}) = \times P(X_{i1})P(X_{i2})} \rightarrow P(X_{i1}|\cancel{X_{i2}}) = P(X_{i1})$

$X_{ij} \perp X_{im}|C_k$

C_k = has Diabetes

blood pressure, weight

$P(X_{ij}|X_{im}, C_k) = P(X_{ij}|C_k)$

diabetes (NB)

$$\underbrace{P(X_{ij}|X_{im},C_k)}$$

- $P(X_i|C_k)P(C_k) = P(X_{i1}|C_k)P(X_{i2}|C_k),\dots,P(X_{ip}|C_k)\times P(C_k)$

$$= P(C_k)\prod_{j=1}^{p} P(X_{ij}|C_k)$$

e_1	b	~~d~~	~~e~~	b	b	~~d~~	~~e~~			
e_2	b	c	e	b	b	d	d	e	c	c
e_3	a	d	a	d	e	a	e	e		
e_4	b	a	d	b	e	d	a	b		
e_5	a	b	a	b	a	b	a	e	d	
e_6	a	c	a	c	a	c	a	e	d	
e_7	e	a	e	d	a	e	a			
e_8	d	e	d	e	d					

e_1–e_4: spam (junk

e_5–e_8: ham/not junk

$k = \{\text{spam } (+), \text{ham} (-)\} = \{1,0\}$

- $P(C_+) = P(C_-) = \frac{1}{2}$ (uniform (class) prior)
- class-conditional dist: choice of model $\rightarrow$ how we represent each email WD.

① Multivariate Bernoulli: does the word appear?

② Multinomial NB: what is the prob. of seeing the word?

$X_{ia} = \{$

	X_{ia}	X_{ib}	X_{ic}
e_1			
e_2	0	1	0
e_3	0	1	1
e_4	1	0	0
e_5	1	1	0
e_6	1	0	1
e_7	1	0	0
e_8	0	0	0

Vocabulary $= V = \{a,b,c\}$ a = "Casino" = "Meeting"

d = 'stop' word = "the", "and" (NLP) (Computer vision)

"Gamble"

$X_{iv} \mid C+ \sim \text{Bernoulli}(P_v^+)$ → prob. of seeing word v in '+' email

$\boxed{x_1, \ldots, x_n \sim \text{Bern}(\theta) \quad \hat{\theta}_{MLE} = \bar{x}}$

$x_{iv} = \begin{cases} 1 & \text{with } P_v^- \\ 0 & \text{with } 1-p_v^+ \end{cases}$

model params : $P_a^+, P_b^+, P_c^+, P_a^-, P_b^-, P_c^-$

Estimates (MLE) :

$P_a^+ = \frac{2}{4}$	$P_b^+ = \frac{3}{4}$	$P_c^+ = \frac{1}{4}$
$P_a^- = \frac{3}{4}$	$P_b^- = \frac{1}{4}$	$P_c^- = \frac{1}{4}$

$e_* = abbdebb \qquad c_* = ?$

$x_* = (x_{*a}, x_{*b}, x_{*c})^T$
$= (1,1,0)^T$

$C_* = \underset{k \in \{+,-\}}{\text{argmax}}\, P(C_k)P(x_* \mid C_k)$

$= \text{argmax}\, \{P(C_+)\underbrace{P(x_* \mid C_+)}, P(C_-)\underbrace{P(x_* \mid C_-)}\}$

$= \text{argmax}\, \{\frac{9}{64}, \frac{9}{128}\}$

$= +$

(Bern NB)

$P(C_+)\underbrace{P(x_* \mid C_+)} = P(C_+)\prod_{v \in V} P(x_v = x_{*v} \mid C_+)$

$= P(C_*)P(X_a = x_{*a} \mid C_+)P(X_b = x_{*b} \mid C_+)$
$\times P(X_c = x_{*c} \mid C_+)$

$X_a \mid C_+ \sim \text{Bern}(P_a^+)$
$P(X_a = 1 \mid C_+)$
$= P_a^+$

$P(C_*)\underbrace{P(X_a = 1 \mid C_+)}\,\underline{P(X_b = 1 \mid C_+)}\,\underline{P(X_c = 0 \mid C_+)}$

$P(C_*) \cdot P_a^+ \cdot P_b^+ \cdot (1 - P_c^+)$

$= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot (1 - \frac{1}{4})$

$= \frac{3}{16} \cdot \frac{3}{4} = \frac{9}{64}$

$$\boxed{\begin{aligned} & P(C_-)P(X_a = x_{*a} \mid C_-) \ldots \\ &= \frac{1}{2} \cdot P_a^- \cdot P_b^- \cdot (1 - P_c^-) \\ &= \frac{9}{128} \end{aligned}}$$

Multinomial NB

$X \sim \text{Multinomial}(n, K, \theta_1, \ldots, \theta_k)$ → dist on histograms (Generalising Binomial to K outcomes)

repeat an experiment n times

and there are K possible outcomes.

$X_1, x_2, \ldots, X_n \sim \text{Multinom}(n, k, \theta_1, \ldots, \theta_k)$, $X_1 \in \mathbb{R}^k$

- Fair dice → $K = 6$, $\theta_1 = \frac{1}{6} = \theta_2 = \theta_3 \ldots = \theta_6$

$X \sim \text{Binom}(n, p)$

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

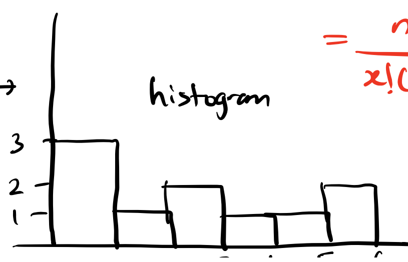
$= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$

- $n = 10$

6, 6, 5, 4, 1, 1, 1, 3, 3, 2 counts

$x = \underbrace{(3, 1, 2, 1, 1, 2)}$ →

histogram



- $P(X = x) = P(X = (x_1, \ldots, x_k)) = \frac{n!}{x_1! \ldots x_k!} \theta_1^{x_1}\theta_2^{x_2} \ldots \theta_k^{x_k}$,

$\sum_{j=1}^{k} \theta_j = 1$ 1 2 3 4 > 0

	x_{ia}	x_{ib}	x_{ic}
e_1	0	3	0
e_2	0	3	3
e_3	~~3~~ 0	0	0
e_4	~~2~~ 0	3	0
e_{p+}	1	1	1
e_5	4	3	0
e_6	4	0	3
e_7	3	0	0
e_8	0	0	0
e_{p-}	1	1	1

$e_{p+} = (1,1,1)$

$X | C_+ \sim \text{Multinomial}(n, k, \theta_1, \dots, \theta_k)$

$n = 17 \quad k = 3 \quad \theta_a^+, \theta_b^+, \theta_c^+$

$X | C_- \sim \text{Multinom}(n = 17, k = 3, \theta_a^-, \theta_b^-, \theta_c^-)$

$$\boxed{\begin{array}{lll} \theta_a^+ = \dfrac{\cancel{3}\,0+1}{17+3} = 0 & \theta_b^+ = \dfrac{9+1}{17+3} & \theta_c^+ = \dfrac{3+1}{17+3} \\ \theta_a^- = \dfrac{11+1}{17+3} & \theta_b^- = \dfrac{3+1}{17+3} & \theta_c^- = \dfrac{3+1}{17+3} \end{array}}$$

$$P(C_+)P(X_* | C_+) = P(C_+) \frac{n!}{x_{*a}!\, x_{*b}!\, x_{*c}!} (\theta_a^+)^{x_{*a}} (\theta_b^+)^{x_{*b}} (\theta_c^+)^{x_{*c}}$$

$$= \frac{1}{2} \cdot \frac{17!}{1!\,4!\,2!} \left(\frac{5}{17}\right)^1 \left(\frac{9}{17}\right)^4 \left(\frac{3}{17}\right)^2$$

$= \boxed{}$

$\frac{1}{15} + \frac{10}{16} + \frac{4}{15} - \frac{15}{15} = 1$

$e_* = abbdebbcc$

$x_* = (1, 4, 2)$