Minimum Risk Classifier

$c \rightarrow$ states of nature $\{\omega_1, \omega_2, \ldots, \omega_c\}$

$a \rightarrow$ action $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$

$x \rightarrow$ d-dimensional feature vector

Loss Function: $\lambda(\alpha_i/\omega_j) \rightarrow$ loss occurred for taking action α_i when the true state of nature is ω_j.

$\lambda(\alpha_1/\omega_1) = 0$ $\quad \lambda(\alpha_1/\omega_2) = 1$

$\lambda(\alpha_2/\omega_1) = 3$ $\quad \lambda(\alpha_2/\omega_2) = 0$

Risk Function
Avg. loss
~~Avg. loss~~
conditional risk
expected loss

$$R(\alpha_1/x) = \sum_{j=1}^{c} \lambda(\alpha_1/\omega_j) P(\omega_j/x)$$

$$\lambda(\alpha_i/\omega_j) = \lambda_{ij}$$

$$R(\alpha_1/x) = \lambda_{11} P(\omega_1/x) + \lambda_{12} P(\omega_2/x)$$

$$R(\alpha_2/x) = \lambda_{21} P(\omega_1/x) + \lambda_{22} P(\omega_2/x)$$

if $\alpha_1/x < \alpha_2/x \Rightarrow$ take ~~the~~ action α_1

then this condition holds true

$$(\lambda_{21} - \lambda_{11}) P(\omega_1/x) > (\lambda_{12} - \lambda_{22}) P(\omega_2/x)$$

Ques An incoming email is either a normal (imp. mail) $[\omega_1]$ or a junk mail $[\omega_2]$ ~~who~~ we have two actions, α_1 (keep the mail) and α_2 (put the mail to trash/null/dev)

Minimum error rate classifier

$$\lambda(\alpha_i/\omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad i, j = 1, \ldots, c$$

$$R(\alpha_i/x) = \sum_{i \neq j} P(\omega_j/x)$$

$$R(\alpha_1/x) = P(\omega_2/x)$$

$$R(\alpha_2/x) = P(\omega_1/x)$$

$$R(\alpha_i/x) = \sum_{i \neq j} P(\omega_j/x) = 1 - P(\omega_i/x)$$

$\alpha_1/\omega_1 = 0$ $\alpha_1 = \text{good}$ $\omega_1 = \text{accept}$

$\alpha_1/\omega_2 = 3$ $\alpha_2 = \text{don't}$ $\omega_2 = \text{reject}$

$\alpha_2/\omega_1 = 1$

$\alpha_2/\omega_2 = 0$

$R(\alpha_1/x) = 1 - P(\omega_1/x)$

$R(\alpha_2/x) = 1 - P(\omega_2/x)$

$\downarrow$ Minimize / Maximize — $\downarrow$ Maximize / Minimize

$P(\omega_1) = 0.4$ $P(\omega_2) = 0.6$

$P(x/\omega_1) = 0.35$ $P(x/\omega_2) = 0.65$

$$R(\alpha_1/x) = 0.1 \left[\frac{0.65 * 0.6}{0.65 * 0.6 + 0.35 * 0.4} \right]$$

$$= 0.73$$

$$R(\alpha_2/x) = 3 \left[\frac{0.35 * 0.4}{0.65 * 0.6 + 0.35 * 0.4} \right] + 0$$

0.53

0.26

$$= 0.79$$

$\alpha_2/x > \alpha_1/x \Rightarrow \alpha_1$

~~Pag~~ P(accept / good) $P(\omega_2/x_1)$

$$= \frac{P_{ga} * P_{accept}}{P_{ga} * P_{accept} + P_{gr} P_{reject}}$$

M-estimate

$$P_{tail} = \frac{N_{tail} + M\,\Pi_{tail}}{N_{all} + m}$$

$M\,\Pi_{tail}$ ← prior expectat.

m ← confidence

— if $N_{all} = N_{tail} = 0$ ⇒ degenerates to prior expectation
if N_{all} & N_{tail} → large ⇒ relative frequency

Toss No.	1	2	3	4	5
outcome	tail	tail	Head	tail	Head
Relative freq.	1.0	1.0	0.67	0.75	0.60
m-estimate	0.66	0.75	0.8	4/6 0.83	4/8
m=100	0.504	0.509	0.507		
	1/1	2/2	2/3	3/4	3/5
m=1	0.75	0.83	0.625		

$\pi_{tail} = 0.5$

$m = 2$

Imp: prior expectation helps us to improve probability estimates in domains with insufficient no. of observations.

4 times toss
3 times appears tails

P(tail) = 0.75

$\frac{1+P}{1+2}$

$P_{pos}(x) = P_{pos}(at_1) \cdot P_{pos}(at_2) \cdot P_{pos}(at_3)$

$P_{neg}(x) = P_{neg}(at_1) \cdot P_{neg}(at_2) \cdot P_{neg}(at_3)$

Gaussian

classmate Date 11/2/19 Page

Ex:	at_1	at_2	at_3	
ex_1	3.1	2.1	2.3	+ve
ex_2	4.2	6.2	7.6	+ve
ex_3	7.8	1.3	0.5	+ve
ex_4	2.3	5.2	2.4	−ve
ex_5	6.4	3.2	4.3	−ve
ex_6	1.3	5.8	3.3	−ve

$$P(x) = k\left(\sum_{i=1}^{m} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\right)$$

$$k = \frac{1}{\sqrt{2\pi\sigma^2}} \quad \frac{1}{\sqrt{(2\pi)^{m/2}}}$$

$$x = (9, 2.5, 3.2)$$

$$P_{pos}(at_1) = \frac{1}{\sqrt{(2\pi)^3}}\left[e^{-0.5(at_1-3.1)^2} + e^{-0.5(at_1-4.2)^2} + e^{-0.5(at_1-7.8)^2}\right]$$

$$= 0.03\left[\cancel{121.113} + \cancel{13.974} + \left[2.76\times10^{-8} + 992.9\times10^{-8}\right.\right.$$

$$P_{neg}(at_1) = \frac{1}{\sqrt{(2\pi)^3}}\left[e^{-0.5(at_1-2.3)^2} + e^{-0.5(at_1-6.4)^2} + e^{-0.5(at_1-1.3)^2}\right]$$

$$= 0.068\left[1.787\times10^{-10} + 0.034 + 4.23\times10^{-13}\right]$$

$$= 0.00234$$

$$P_{pos}(at_2) = \frac{1}{\sqrt{(2\pi)^3}}\left[e^{-0.5(at_2-2.1)^2} + e^{-0.5(at_2-6.2)^2} + e^{-0.5(at_2-1.3)^2}\right]$$

$$= 0.088$$

$\cancel{17.405}$ $\cancel{11.52}$

$$P_{neg}(at_2) = \frac{1}{\sqrt{(2\pi)^3}}\left[e^{-0.5(at_2-5.2)^2} + e^{-0.5(at_2-3.2)^2} + e^{-0.5(at_2-5.8)^2}\right]$$

−3.645 −0.245 −5.445

$$= 0.063\left[0.0261 + 0.782 + 0.0043\right]$$

$$= 0.063\left[0.8134\right]$$

$$= 0.0512$$

$$P_{pos}(at_3) = 0.063\left[e^{-0.5(at_3-2.3)} + e^{-0.5(at_3-7.6)} + e^{-0.5(at_3-0.5)}\right]$$

$$= 0.063 \times 0.6931 = 0.04366$$

$$P_{neg}(at_3) = 0.063\left(e^{-0.5(at_3-2.4)} + e^{-0.5(at_3-4.3)} + e^{-0.5(at_3-3.3)}\right)$$

$$= 0.063(0.726 + 0.546 + 0.995)$$

Date
Page

$P_{pos}(x) = 0.0001152$

$P_{neg}(x) = 0.00000156$

14/2/19

Maximum Likelihood Estimation

YYY NNNNN

What is priori probability of Y that maximizes this likelihood of data P(Y).

We denote prob. by symbol π, i.e., $P(Y) = \pi$

incoming $P(Y_i) = \begin{cases} \pi & \text{if } Y_i = Y \\ 1-\pi & \text{if } Y_i = N \\ \cancel{Y_i \; \pi \; 1-Y_i} \end{cases}$

111 00000

i.e., $P(Y_i) = \pi^{Y_i}(1-\pi)^{1-Y_i}$

if Y_i becomes 1 than $(1-\pi)^{1-Y_i}$ becomes 1

if Y_i becomes 0 π^{Y_i} becomes 1

Assuming data independence, then joint probability

$$P(\text{data}) = \prod_{i=1}^{n} P(Y_i)$$

$$= \pi^{\text{count}(Y_i=1)} (1-\pi)^{\text{count}(Y_i=0)}$$

$$= \pi^3 (1-\pi)^5$$

Find π that maximizes the expression, i.e., maximizes the logarithm of the expression

$$\log P(\text{data}) = 3 \cdot \log \pi + 5 \log(1-\pi)$$

$$\frac{\nabla \cdot \log P(\text{data})}{\nabla \cdot \pi} = 0$$

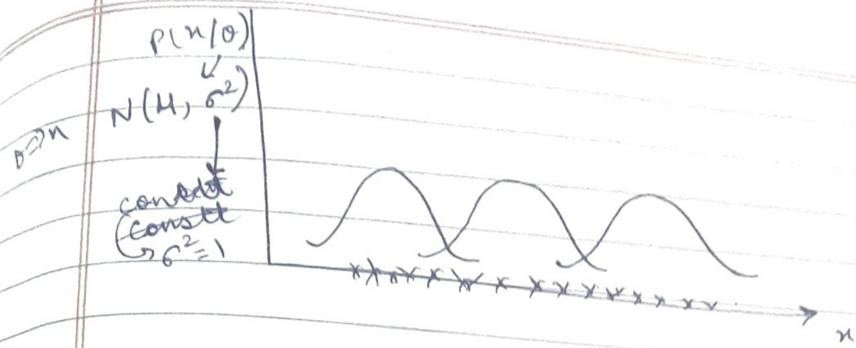
$$\Rightarrow \frac{3}{\pi} - \frac{5}{(1-\pi)} = 0$$

$$\Rightarrow \frac{3}{\pi} = \frac{5}{1-\pi}$$

$$\Rightarrow 3(1-\pi) = 5\pi$$

$$\Rightarrow \pi = \frac{3}{8}$$

$\theta \Rightarrow$ Parameter



$$\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = 0$$

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} * 2(x-\mu) = 0$$

$$e^{-(x-\mu)^2/2} (x-\mu) = 0$$

$$P(D/\theta) = \prod_{i=1}^{K} P(x_k/\theta)$$

15/2/19

Nearest Neighbour Classifier (used for post processing)

Ex	Shape	Size	Volume	Colour	Shade	Class	#diff
x	Unknown data					?	
Ex1	Training data					Yes	3
Ex2						Yes	4
Ex3						Yes	4
Ex4						Yes	1
Ex5						No	2
Ex6						No	3
Ex7						No	5
Ex8						No	5
Ex10						No	6
Ex10						No	

always take K to be odd ← K nearest neighbour → we take min. K-values

but taking odd no. of K in multiclass problem

$$d_E(x,y) = \sqrt{\sum_{i=1} (x_i - y_i)^2}$$

$x = [1, 4, 3]$

$ex_1 \{[1, 3, 2], poo\}$

$ex_2 \{[3, 5, 2], poo\}$

$ex_3 \{[2, 3, 5], neg\}$

$ex_4 \{[5, 2, 4], neg\}$

Normalising [0:1]

$$x = \frac{x - MIN}{MAX - MIN}$$

$[7, 4, 25, -1, 10]$

$x = (1, 0.2, 284)$

$y = (0, 0.8, 162)$

there is class of which these belongs

training data

one data is dominating

so we have to normal that value by using formula.

i.e, $x = \frac{x - Min}{Max - Min}$

$[7, 4, 25, -1, 10]$

$$x = \frac{x - Min}{Max - Min}$$

So, $x - Min = \frac{7 - (-1)}{25 - (-1)} = \frac{8}{26} = 0$

$[7, 4, 25, -1, 10]$

$7 \rightarrow 0$

$4 \rightarrow \frac{4 - (-1)}{25 - (-1)} = \frac{5}{26}$

$25 \rightarrow \frac{25 - (-1)}{25 - (-1)} = 1$

$-1 \rightarrow \frac{-1 - (-1)}{25 - (-1)} = 0$

$10 \rightarrow \frac{10 - (-1)}{25 - (-1)} = \frac{11}{26}$

$[0.307, 0.19, 1, 0, 0.42]$

$\rightarrow$ Normalized

Midsem syll till here

Weighted Nearest Neighbour (Do on own)

$N = 21$

$P(0-10) = 1/21$

$P(11-20) = 2/21$

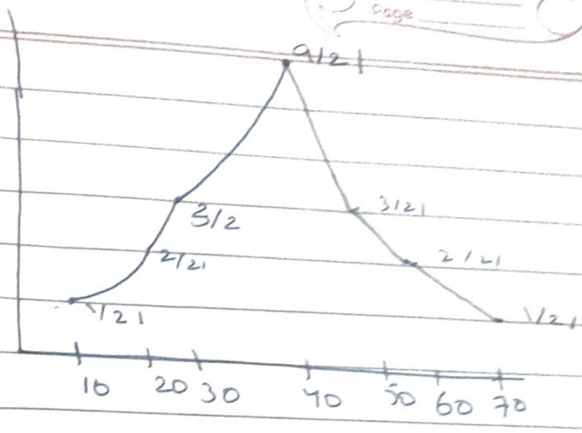
$P(21-30) = 3/21$

$P(31-40) = 9/21$

$P(41-50) = 3/21$

$P(51-60) = 2/21$

$P(61-70) = 1/21$



In Approximating Pdfs

we start separate the data and then reform it or combine it.

Suppose,

In prev. example, parents are to be given male/female then we split our data.

Combine Gaussian Function

$$P(x) = k \sum_{i=1}^{m} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$k = \frac{1}{(2\pi)^{m/2} \sigma^m}$$

Note :- Denote μ_i) the value of x in the i^{th} example.