Date
Page

Name: ROHIT GARLA NetID: rg3365

HOMEWORK 2.

1)
(a) True Positive RATE = $\frac{\text{\# True Positive Examples}}{\text{\# of examples whose correct label is}}$

$$= \frac{\#TP}{\#TP + \#FN}$$

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

$$= 0$$

(b) False Positive Rate = $\frac{\#FP}{\#FP + \#TN}$

$$= \frac{1}{1+2} = \frac{1}{3}$$

(c)

$$\text{accuracy} = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN}$$

$$= \frac{56+41}{56+41+2+1} = \frac{97}{100} = 0.97$$

Date
Page

2)

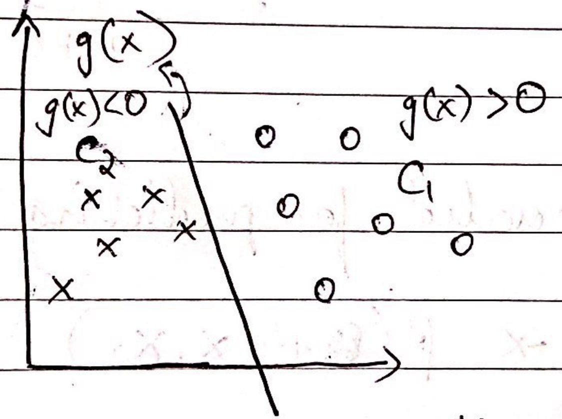
Answer) Now for the two classes we can take,

$g(x) = 5x_2 + 3x_1 - 4$ for Class C_1

$g_2(x) = -3x_2 + 2x_1 - 6$ for class C_2

Now, we can generate $g(x)$ as.

$$g(x) = \begin{cases} C_1 & \text{if } g(x) > 0 \\ C_2 & \text{otherwise.} \end{cases}$$



Since g(x) will be line that seperates two classes.

$$g(x) = g_1(x) - g_2(x)$$
$$= (5x_2 + 3x_1 - 4) - (-3x_2 + 2x - 6)$$
$$= 5x_2 + 3x_1 - 4 + 3x_2 - 2x + 6.$$
$$= 8x_2 + x_1 + 2.$$

$w_2 = 8$; $w_1 = 1$; $w_0 = 2$

A 3) $P(Spam/x_1 x_2) = \frac{1}{1+e^{-(3x_2 - 2x_1 + 1)}}$

given in $x_2 = 2$, $x_1 = 3$.

$P(Spam/x_1 x_2) = \frac{1}{1+e^{-(6-6+)}} = \frac{e}{1+e}$.

$P(Not\ Spam/x_1 x_2) = 1 - P(Spam)$

$= 1 - \frac{e}{1+e}$

$= \frac{1}{1+e}$

Cost associated for predicting Not Spam

$= a \times P(+/x_1 x_2)$

$= a \times \frac{e}{1+e} = 5 \times \frac{e}{1+e} = \frac{5e}{1+e}$.

Cost associated for predicting spam

$= b \times P(-/x_1 x_2)$

$= 2 \times \frac{1}{1+e} = \frac{2}{1+e}$.

The Spam classification has lower risk.

$$\frac{2}{1+e} < \frac{5e}{1+e}$$

A4)
(a) Bias = ~~Exp~~ Estimate − actual

$$\text{Estimate} = \frac{\sum_{x \in X} x}{N+1}$$

$$\text{Actual} = \frac{\sum_{x \in X} x}{N}$$

$$\frac{\sum_{x \in X} x}{N+1} - \frac{\sum_{x \in X} x}{N} = \text{Bias}$$

$$\frac{N\left(\sum_{x \in X} x\right) - \left[N\left(\sum_{x \in X} x\right) + \left(\sum_{x \in X} x\right)\right]}{(N+1)N} = \text{Bias}$$

$$-\frac{\sum_{x \in X} x}{(N+1)N} = \text{Bias}$$

(b) Even for D which is an exponential distribution the answer wouldn't change

Bias would still be estimate − actual

So bias would not change

MLE estimate with $\hat{\lambda} = \text{argmin}\, L(\lambda, x)$ $\Rightarrow \hat{\lambda} = 1/\bar{x}$

where $\bar{x}$ is mean.

(A5)

For class +

$$\hat{\mu} = \frac{\sum_{t=1}^{N} x^{t}}{N}$$

$$\mu^{+}_{x_1} = \frac{2.7 + 3.2 - 0.4}{3} = 1.833$$

$$\mu^{+}_{x_2} = \cancel{\frac{0.6+1.8+2.1}{3}} = \frac{4.8 + 5.1 - 0.3}{3}$$

$$= 3.2$$

For class −

$$\mu^{-}_{x_1} = \frac{0.6 + 1.8 + 2.1}{3} = 1.5$$

$$\mu^{-}_{x_2} = \frac{0.5 + 2.8 + 4.3}{3} = 2.53$$

$$\Sigma_{+} = \frac{1}{N} \begin{bmatrix} x^{+}_{11} - m^{+}_{x_1} & x^{+}_{12} - m^{+}_{x_1} & x^{+}_{13} - m^{+}_{x_1} \\ x^{+}_{21} - m^{2}_{x} & x^{+}_{22} - m^{L}_{x_1} & x^{+}_{23} - m^{+}_{x_1} \end{bmatrix} \times \begin{bmatrix} x^{+}_{11} - m^{+}_{x_1} & m^{+}_{21} - m^{+}_{x_2} \\ x^{+}_{12} - m^{+}_{x_1} & m_{22} - m^{+}_{x_1} \\ x^{+}_{13} - m_{x_1} & x^{+}_{23} - m^{+}_{x_1} \end{bmatrix}$$

$N = 3$

$$= \frac{1}{3} \begin{bmatrix} 27-21.833 & 3.2-1.833 & -(-0.4)-1.833 \\ 4.8-3.2 & 5.1-3.2 & -(0.3)-3.2 \end{bmatrix}$$

$$* \begin{bmatrix} 27-1.833 & 4.8-3.2 \\ 3.2-1.833 & 5.1-3.2 \\ -0.4-1.833 & -(0.3)-3.2 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 27-21.833 & 3.2-1.8333 & -(-0.4)-1.833 \\ 4.8-3.2 & 5.1-3.2 & -(0.3)-3.2 \end{bmatrix}$$

$$* \begin{bmatrix} 0.867 & 1.6 \\ 1.367 & 1.9 \\ -2.233 & -3.5 \end{bmatrix}$$

$$= \begin{bmatrix} 2.535 & 3.933 \\ 3.933 & 6.14 \end{bmatrix}$$

$$\Sigma_{-} = \frac{1}{3} \begin{bmatrix} 0.6-1.5 & 1.8-1.5 & 2.1-1.5 \\ 0.5-2.5333 & 2.8-2.5333 & 4.3-2.5333 \end{bmatrix} \begin{bmatrix} 0.6-2.5 & 0.5-2.533 \\ 1.8-1.5 & 2.8-2.533 \\ 2.1-1.5 & 4.3-2.533 \end{bmatrix}$$

$$= \begin{bmatrix} 0.42 & 0.99 \\ 0.99 & 2.4422 \end{bmatrix}$$

(5b)

A)

$$\log p(x/+) = \log\left[\frac{1}{(2\pi)^{d/2}|\Sigma_+|^{1/2}} e^{-1/2 (x-\mu)^T \Sigma^{-1}(x-\mu)}\right]$$

$d = 2$

$$|\Sigma_+| = \begin{vmatrix} 2.535 & 3.933 \\ 3.933 & 6.14 \end{vmatrix} = 0.096411$$

$$\Sigma_+^{-1} = \begin{bmatrix} 63.6857 & -40.794 \\ -40.794 & 26.2937 \end{bmatrix}$$

$$x = \begin{bmatrix} 1.6 \\ 2.3 \end{bmatrix} \quad \mu_+ = \begin{bmatrix} 1.833 \\ 3.2 \end{bmatrix}$$

$$\log p(x|+) = \log e^{-1/2\left[\begin{bmatrix} 1.6-1.833 & 2.3-3.2 \end{bmatrix}\begin{bmatrix} 63.6857 & -40.794 \\ -40794 & 26.2937 \end{bmatrix}\begin{bmatrix} 1.6-1.83 \\ 2.3-3.2 \end{bmatrix}\right]}$$

$$= \quad - \log\left((2\times 3.14)^{2/2} \times (0.096411)^{1/2}\right)$$

$$= \log_e e^{-1/2}\left[7.646\right] - \cancel{1.0158} - 1.95093$$

$$= \cancel{0.41719}$$

$$\cancel{= \log_e e^{-1/2}\left[7.646\right] - 1.0581}$$

$$\cancel{-4.8811} = -5.77406$$

$$\left[\Sigma \right] = \begin{bmatrix} 0.42 & 0.99 \\ 0.99 & 2.4422 \end{bmatrix}$$

$|\Sigma| = 0.04512$

$$\left[\Sigma \right] = \begin{bmatrix} 0.42 & 0.99 \\ 0.99 & 2.4422 \end{bmatrix} \quad \cancel{0.03572}$$

$$\Sigma^{-1} = \begin{bmatrix} 53.5288 & -21.699 \\ -21.699 & 9.205 \end{bmatrix}$$

$$x = \begin{bmatrix} 1.6 \\ 2.3 \end{bmatrix} \quad \mu = \begin{bmatrix} 1.5 \\ 2.533 \end{bmatrix}$$

$$\log p(x|-) = \log e^{-1/2 \left[\begin{bmatrix} 1.6-1.5 & 2.3-2.533 \end{bmatrix} \begin{bmatrix} 53.5285 & -21.699 \\ -21.699 & 9.205 \end{bmatrix} \begin{bmatrix} 1.6-1.5 \\ 2.3-2.583 \end{bmatrix} \right]} - \log\left((2\times 3.14)^{2/2} \times (0.04572)^{1/2} \right)$$

$= \log(e^{-1/2}(2.046191)) - \log(1.34207)$

$= \frac{-1}{2}(2.046191) - \log(1.34207)$

$= -1.3$

using ML hypothesis

we can conclude,

$$\underset{\Theta}{\operatorname{argmax}}\left[\log P(x|+), \log P(x|-)\right]$$

$\Rightarrow$ -ve class is the predicted class.

$$\log p(x/-) = -1.3$$

$$\log p(x|+) = -5.77406$$

$|\Sigma_+| = 0.096411$

$|\Sigma_-| = 0.045624$.

$\therefore \log P(x/-) > \log P(x/+)$

It belongs to negative class.

8(c)

A)(i) Under some circumstances, the attributes do not have much relation with tokens, therefore distribution or variance may be very similar. We can consider them as one for example: like ";" tokens in the reviews.

(ii) However, there are many other situations that covariant matrix differs a lot under different labels. For example: Frequency of positive words in a single review may have more impact

2)(a) In linear regression, we have a linear model

$$g(x^t | w_1, w_0) = w_1 x^t + w_0$$

And after taking the derivative of sum of squares error w.r.t w_1 and w_0, we get,

$$\sum_t r^t = N w_0 + w_1 \sum_t x^t$$

$$\sum_t r^t x^t = w_0 \sum_t x_t + w_1 \sum_t (x^t)^2.$$

which can be written in vector-matrix form

as $Aw = y$ where:

$$A = \begin{bmatrix} N & \sum_t x^t \\ \sum_t x^t & \sum_t (x^t)^2 \end{bmatrix}, \quad w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \quad y = \begin{bmatrix} \sum_t r^t \\ \sum_t r^t x^t \end{bmatrix}$$

can be solved as $w = A^{-1} y$

$N = 10$

$$A = \begin{bmatrix} 10 & 7221 \\ 7221 & 6500115 \end{bmatrix} \qquad Y = \begin{bmatrix} 527 \\ 477908 \end{bmatrix}$$

$$|A| = 12858309$$

We know for 2x2 matrix,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \underbrace{\frac{1}{ad - bc}}_{\text{determinant}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

So, for A we get

$$A^{-1} = \frac{1}{|A|} \times \begin{bmatrix} 6500115 & -7221 \\ -7221 & 10 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} \dfrac{6500115}{12858309} & \dfrac{-7221}{12858309} \\[2ex] \dfrac{-7221}{12858309} & \dfrac{10}{12858309} \end{bmatrix}$$

$$\cancel{A^{-1} = \begin{bmatrix} 0.5055 & -0.000562 \\ \text{[illegible]} & 0.000001 \end{bmatrix}}$$

$$W = A^{-1} Y$$

$$= \begin{bmatrix} \frac{6500115}{12858309} & \frac{-7221}{12858309} \\ \frac{-7221}{12858309} & \frac{10}{1285309} \end{bmatrix} \begin{bmatrix} 527 \\ 477908 \end{bmatrix}$$

$$W = \begin{bmatrix} -\frac{8471021}{4286103} \\ \frac{973613}{12858309} \end{bmatrix}$$

$$w_0 = \frac{-8471021}{4286103} = -1.9763$$

$$w_1 = \frac{973613}{12858309} = 0.075719$$

$$g(x) = (0.075719)x - 1.9763.$$

(7) $\because x = 475$

$$g(x) = (0.075719) \times 475 - 1.9763$$

$$= 35.9665 - 1.9763$$

$$= 33.99$$

$$\approx 34 \text{ stories}$$

7) a)

Original dataset :

x_1	x_2	Label
2.5	42	+
3.8	51	+
-0.3	-1	+
0.7	3	-
1.6	26	-
2.3	41	-

Scaling formula = $\dfrac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}$.

Scaled dataset :

x_1	x_2	label
0.68	0.83	+
1.00	1.00	+
0.00	0	+
0.24	0.08	-
0.46	0.52	-
0.63	0.80	-

$$x_{21}(\text{scaled}) = \frac{42-(-1)}{51-(-1)} = \frac{43}{52} = 0.83$$

$$x_{22} = \frac{51-(-1)}{51-(-1)} = \frac{52}{52} = 1$$

$$x_{23} = \frac{-1-(-1)}{51-(-1)} = 0$$

$$x_{24} = \frac{3-(-1)}{51-(-1)} = 0.08$$

$$x_{25} = \frac{26-(-1)}{51-(-1)} = \frac{27}{52} = 0.52$$

$$x_{26} = \frac{41-(-1)}{51-(-1)} = \frac{42}{52} = 0.80$$

1)(b) Test data $x = \begin{bmatrix} 3.9 \\ 4 \end{bmatrix} \Rightarrow x_1 = 3.9$, $x_2 = 4$.

Scaling formula : $\dfrac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}$

Scaled Test data ; $x = \begin{bmatrix} 1.02 \\ 0.096 \end{bmatrix}$

$$x_1 = \frac{3.9-(-0.3)}{3.8-(-0.3)} = \frac{4.2}{4.1} = 1.02$$

$$x_2 = \frac{4-(-1)}{51-(-1)} = \frac{5}{52} = 0.096$$

Date
Page

Now we calculate the Euclidian distance to all the scaled attributes to (1.02, 0.096)

For class = + ;

$(X_1, X_2) = (0.68, 0.83)$; ~~class = +~~

$$E.D = \sqrt{(0.68-1.02)^2+(0.096-0.83)^2}$$

$$= 0.808$$

$(X_1, X_2) = (1, 1)$; ~~class = +~~

$$E.D = \sqrt{(1-1.02)^2+(1-0.096)^2} = 0.904$$

$(X_1, X_2) = (0, 0)$

$$E.D = \sqrt{(1.02)^2+(0.096)^2} = 1.0245$$

~~(X_1, X_2)~~ For class = − ;

$(X_1, X_2) = (0.24, 0.08)$

$$ED = \sqrt{(1.02-0.24)^2+(0.096-0.08)^2}$$

$$= 0.78$$

$(X_1, X_2) = (0.46, 0.52)$

$$ED = 0.702$$

$(x_1, x_2) = (0.63, 0.80)$

$$ED = \sqrt{(0.63-1.02)^2 + (0.80-0.096)^2}$$
$$= 0.804$$

From the above results, we can see that

ED is lowest when $(x_1, x_2) = (0.46, 0.52)$

So the predicted label is "—"