CSC 5521 – Intro to Machine Learning

HW1

Saksham Goel

goelx029 | 5138568

CSCI 5521
Homework 1

Name - Saksham Goel
ID - goelx029 | 5138568

Q1 Let - D = Event that a person has the disease
+ = Event that the checkup comes '+' for disease
− = " " " " " '−' for disease

Given:

$P(D) = 1/10000$ $\quad$ $P(+|D) = 0.983$

$P(\bar{D}) = 9999/10000$ $\quad$ $P(-|D) = 0.017$

know $P(+|D) + P(-|D) = 1$

$P(+|\bar{D}) = 0.055$

$P(-|\bar{D}) = 0.945$

know $P(+|\bar{D}) + P(-|\bar{D}) = 1$

Need to find

a) $$P(D|+) = \frac{P(+|D)\,P(D)}{P(+|D)P(D) + P(+|\bar{D})\,P(\bar{D})} = \frac{0.983 \times \cancel{10^{-4}}}{0.983 \times \cancel{10^{-4}} + 0.055 \times 9999 \times \cancel{10^{-4}}}$$

$$= \frac{0.983}{0.983 + 549.945} = \frac{0.983}{550.928} \approx 0.00178$$

$\boxed{0.178\%}$ chances

b) $C(T|\bar{D}) = 1000$

$C(\bar{T}|D) = 1000000$

$C(T|D) = 0$

$C(\bar{T}|\bar{D}) = 0$

Expected Cost of Treatment (when no test done) $= C(T|\bar{D}) \times P(\bar{D}) + C(T|D)\,P(D)$

$= 1000 \times 0.9999 + 0 \times 0.0001$

$= \boxed{\$999.9}$

$E(C(\bar{T})) = C(\bar{T}|\bar{D})\,P(\bar{D}) + C(\bar{T}|D)\,P(D) = 1000000 \times \frac{1}{10000} = \boxed{100}$

c) $E(T|+)$

$\downarrow$

Expected cost given + test

$$E(T|+) = E(T|D)\,P(D|+) + E(T|\bar{D})\,P(\bar{D}|+)$$
$$= 0 \times 0.00178 + 1000 \times 0.99822$$
$$= \boxed{\$998.22}$$

d) We know

$E(T) = \$999.9$

$E(T|+) = \$998.22$

finding

$$E(\tilde{T}|+) = E(\tilde{T}|D)\,P(D|+) + E(\tilde{T}|\bar{D})\,P(\bar{D}|+)$$
$$= 1000000 \times 0.00178 + 0 \times 0.99822$$
$$= \$1780$$

So as $E(T|+)$ is the least $= \$998.22$, we can say going for a treatment when tested +ve for the disease is the least risky option.

e) $E(\bar{T}|-) = E(\bar{T}|D)\,P(D|-) + E(\bar{T}|\bar{D})\,P(\bar{D}|-)$

using Bayes theorem to find

$$P(D|-) = \frac{P(-|D)P(D)}{P(-|D)P(D) + P(-|\bar{D})P(\bar{D})} = \frac{0.017 \times \cancel{10^{-4}}}{0.017 \times \cancel{10^{-4}} + 0.945 \times 9999 \times \cancel{10^{-4}}} = \frac{0.017}{9449.072}$$
$$= 1.8 \times 10^{-6}$$

$\Rightarrow E(\bar{T}|-) = 1000000 \times 1.8 \times 10^{-6} + 0 \times (1 - 1.8 \times 10^{-6})$

$\boxed{E(\bar{T}|-) = \$1.8}$

Q2 a) given $P(x|C_1) = \begin{cases} 1/4 & 0 \le x < 4 \\ 0 & \text{otherwise} \end{cases}$

$$P(x|C_2) = \begin{cases} x-1 & 1 \le x < 2 \\ 3-x & 2 \le x < 3 \\ 0 & \text{other} \end{cases}$$

$$P(C_1) = P(C_2) = 0.5$$

then

$$P(C_1|x=1.5) = \frac{P(x=1.5|C_1) \times \cancel{P(C_1)}}{P(x=1.5|C_1)\cancel{P(C_1)} + P(x=1.5|C_2)\cancel{P(C_2)}} = \frac{1/4}{1/4 + 1/2} = \boxed{\frac{1}{3}}$$

$$P(C_2|x=1.5) = \frac{P(x=1.5|C_2)\cancel{P(C_2)}}{P(x=1.5|C_1)\cancel{P(C_1)} + P(x=1.5|C_2)\cancel{P(C_2)}} = \frac{1/2}{1/4 + 1/2} = \boxed{\frac{2}{3}}$$

As $P(C_2|x=1.5) > P(C_1|x=1.5)$ $\boxed{\text{class} = 2}$

or $\log\left(\frac{P(C_1|x)}{P(C_2|x)}\right) = \log\left(\frac{1/3}{2/3}\right) = \log\left(\frac{1}{2}\right) < 0$ hence $\boxed{C_2}$

b) $$P(C_1|x=1.5) = \frac{P(x=1.5|C_1)P(C_1)}{P(x=1.5|C_1)P(C_1) + P(x=1.5|C_2)P(C_2)} = \frac{\frac{1}{\cancel{4}} \times \frac{3}{\cancel{4}}}{\frac{1}{\cancel{4}} \times \frac{3}{\cancel{4}} + \frac{2}{\cancel{4}} \times \frac{1}{\cancel{4}}} = \boxed{\frac{3}{5}}$$

$$P(C_2|x=1.5) = \frac{P(x=1.5|C_2)P(C_2)}{P(x=1.5|C_1)P(C_1) + P(x=1.5|C_2)P(C_2)} = \frac{\frac{2}{\cancel{4}} \times \frac{1}{\cancel{4}}}{\frac{1}{\cancel{4}} \times \frac{3}{\cancel{4}} + \frac{2}{\cancel{4}} \times \frac{1}{\cancel{4}}} = \boxed{\frac{2}{5}}$$

as $\log\left(\frac{P(C_1|x)}{P(C_2|x)}\right) = \log\left(\frac{3/5}{2/5}\right) = \log\left(\frac{3}{2}\right) > 0$ hence

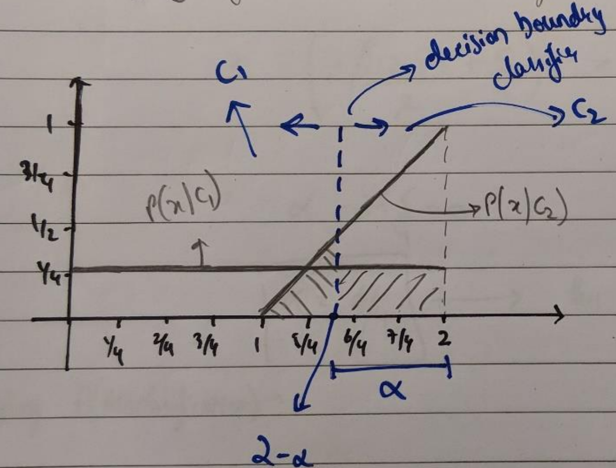
C_1 is the class

c) Our new classifier is given by the following definition:

$$x \longrightarrow \begin{cases} C_2 & \phi(x) < 0 \\ C_1 & \phi(x) \geq 0 \end{cases} \quad \text{where} \quad \phi(x) = |x-2| - \alpha$$

this leads to the following definition → C_2 if $2-\alpha < x < 2+\alpha$ else C_1 for $\alpha \in [0,1]$

drawing this on graph using the symmetry of given distribution & classifier



Given this classifier the probability of misclassification can be given by :⇒

Probability of an object of C_1 with $x > 2-\alpha$
+ Pr " " " " C_2 with $x < 2-\alpha$

↳ equivalent to finding the area covered by the probability density function of each class in the region where misclassified.

$P(\text{misclassification}) = P(\text{misclassifying } C_2 \text{ as } C_1) + P(\text{misclas. } C_1 \text{ as } C_2)$

$$= \left\{ \int_1^{2-\alpha} P(x|C_2)P(C_1)\,dx + \int_{2-\alpha}^{2} P(x|C_1)P(C_2)\,dx \right\} \times 2 \quad (\text{account for symmetry})$$

$$= \left\{ \int_1^{2-\alpha} (x-1)\,dx + \int_{2-\alpha}^{2} \tfrac{1}{4}\,dx \right\} \frac{\times\cancel{2}}{\cancel{2}} = \left\{ \left[\frac{x^2}{2} - x\right]_1^{2-\alpha} + \left[\frac{1}{4}x\right]_{2-\alpha}^{2} \right\}$$

$$P(\text{Misclassification}) = \left\{ \left\{ \frac{(2-\alpha)^2}{2} - (2-\alpha) - \frac{1}{2} + 1 \right\} + \left\{ \frac{1}{4}(2 - 2 + \alpha) \right\} \right\}$$

$$= \left(\frac{(2-\alpha)^2}{2} + \frac{5}{4}\alpha - \frac{3}{2} \right)$$

Now Minimizing $P(\text{Misclass}) \rightarrow \min_{\alpha} P(\text{misclass})$

$$\frac{dP(\text{Miscl})}{d\alpha} = \frac{d}{d\alpha}\left(\left(\frac{(2-\alpha)^2}{2} + \frac{5}{4}\alpha - \frac{3}{2} \right) \right) = 0$$

$$= \left(-\frac{\cancel{2}(2-\alpha)}{\cancel{2}} + \frac{5}{4} \right) = 0$$

$$= \alpha - 2 + \frac{5}{4} = 0$$

$\boxed{\alpha = 3/4} \longrightarrow$ this gives us our optimal boundary.

Finding $P(\text{Misclassification}) \rightarrow$

$$P(\text{Miscl}) = \left\{ \frac{(2-\alpha)^2}{2} + \frac{5}{4}\alpha - \frac{3}{2} \right\} = \left\{ \frac{(2-3/4)^2}{2} + \frac{5}{4} \times \frac{3}{4} - \frac{3}{2} \right\}$$

$$= \left(\frac{25}{\cancel{32}\,16} + \frac{15}{\cancel{16}\,8} - \frac{3}{\cancel{2}} \right) \times \cancel{2} = \frac{25+30}{16} - 3$$

$$= \frac{55-48}{16} = \boxed{\frac{7}{32}}$$

d) Cost ↘ True

		C_1	C_2
classified ~~Actual~~	C_1	-5	$+1$
	C_2	$+3$	-5

~~Probab ↘ True C_1 C_2 classify C_1 C_2~~

Expected cost $\rightarrow$ $-5\times P(\text{correct classify } C_1) - 5\times P(\text{correct classify } C_2)$
$+ 3\times P(\text{Wrong classify } C_1 \text{ as } C_2) + 1\times P(\text{wrong classify } C_2 \text{ as } C_1)$

$$E(\text{cost}) = \Big\{ -5\times \int_0^{2-\alpha} P(x|C_1)\cancel{P(C_1)}\,dx - 5\times \int_{2-\alpha}^{2} P(x|C_2)\cancel{P(C_2)}\,dx$$
$$+ 3\times \int_{2-\alpha}^{2} P(x|C_1)\cancel{P(C_1)}\,dx + 1\times \int_{1}^{2-\alpha} P(x|C_2)\cancel{P(C_2)}\,dx \Big\} \cancel{\times 2}$$

symmetry

$$= -5\int_0^{2-\alpha} \tfrac{1}{4}\,dx - 5\int_{2-\alpha}^{2}(x-1)\,dx + 3\int_{2-\alpha}^{2}\tfrac{1}{4}\,dx + 1\int_{1}^{2-\alpha}(x-1)\,dx$$

$$= -\frac{5}{4}(2-\alpha) - 5\left((2-\alpha) - \frac{(2-\alpha)^2}{2}\right) + \frac{3}{4}\alpha + \left(\frac{(2-\alpha)^2}{2} - (2-\alpha) - \frac{1}{2} + 1\right)$$

$$= -\frac{5}{4}(2-\alpha) - 5(2-\alpha) + \frac{5(2-\alpha)^2}{2} + \frac{3}{4}\alpha + \frac{(2-\alpha)^2}{2} - (2-\alpha) + \frac{1}{2}$$

$$= (2-\alpha)\left(-\frac{5}{4} - 5 - 1\right) + (2-\alpha)^2\left(\frac{5}{2} + \frac{1}{2}\right) + \frac{3}{4}\alpha + \frac{1}{2}$$

$$= 3(2-\alpha)^2 - \frac{29}{4}(2-\alpha) + \frac{3}{4}\alpha + \frac{1}{2} = \boxed{3(2-\alpha)^2 + 8\alpha - 14}$$

finding $\min_{\alpha} E(\text{cost}) \Rightarrow \dfrac{d}{d\alpha}E(\text{cost}) = 0$

$$\frac{d}{d\alpha}\left(3(2-\alpha)^2 + 8\alpha - 14\right) = 0$$

$\Rightarrow -6(2-\alpha) + 8 = 0$

$\Rightarrow -12 + 6\alpha + 8 = 0$

$\boxed{\alpha = 4/6}$ $\boxed{\Rightarrow \alpha = 2/3}$

e) $E(x|c_1) = \int_0^4 x \, P(x|c_1) \, dx$

$= \int_0^4 x \frac{1}{4} dx = \int_0^4 \frac{x}{4} dx = \left[\frac{x^2}{8}\right]_0^4 = \boxed{2}$

$E(x|c_2) = \int_0^4 x \, P(x|c_2) \, dx$

$= \int_1^2 x \, P(x|c_2) \, dx + \int_2^3 x \, P(x|c_2) \, dx$

$= \int_1^2 (x^2 - x) dx + \int_2^3 (3x - x^2) dx$

$= \left[\frac{x^3}{3} - \frac{x^2}{2}\right]_1^2 + \left[\frac{3}{2}x^2 - \frac{x^3}{3}\right]_2^3$

$= \left[\frac{8}{3} - \frac{4}{2} - \frac{1}{3} + \frac{1}{2}\right] + \left[\frac{27}{2} - \frac{27}{3} - \frac{12}{2} + \frac{8}{3}\right]$

$= \left[\frac{7}{3} - \frac{3}{2}\right] + \left[\frac{15}{2} - \frac{19}{3}\right] = \frac{\cancel{12}^{6}}{\cancel{2}} - \frac{\cancel{12}^{4}}{\cancel{3}} = \boxed{2}$

for $Var(x|c_1)$ first $E((x|c_1)^2)$ first

$E((x|c_1)^2) = \int_0^4 x^2 \frac{1}{4} dx = \left[\frac{x^3}{12}\right]_0^4 = \frac{64}{12}$

$Var(x|c_1) = E((x|c_1)^2) - (E(x|c_1))^2 = \frac{64}{12} - 2^2 = \frac{64-48}{12} = \boxed{\frac{16}{12}}$

f. $Var(x|C_2)$ finding $E((x|C_2)^2)$

$$E((x|C_2)^2) = \int_1^2 x^2(x-1)\,dx + \int_2^3 x^2(3-x)\,dx$$

$$= \int_1^2 (x^3-x^2)\,dx + \int_2^3 (3x^2-x^3)\,dx = \left[\frac{x^4}{4}-\frac{x^3}{3}\right]_1^2 + \left[x^3-\frac{x^4}{4}\right]_2^3$$

$$= \left[\frac{16}{4}-\frac{8}{3}-\frac{1}{4}+\frac{1}{3}\right] + \left[27-\frac{81}{4}-8+\frac{16}{4}\right]$$

$$\frac{15}{4}-\frac{7}{3}+19-\frac{65}{4} = 19-\frac{7}{3}-\frac{50}{4}$$

$$= \frac{228-28-150}{12} = \boxed{\frac{50}{12}}$$

$$Var(x|C_2) = E((x|C_2)^2) - (E(x|C_2))^2$$

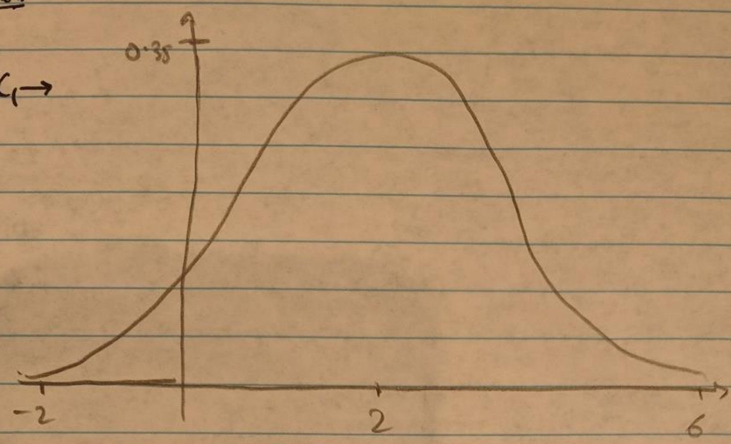
$$= \frac{50}{12} - 4 = \frac{50-48}{12} = \boxed{\frac{2}{12}}$$

Standard Deviation (x | C1) = sqrt(4/3) = 1.1547

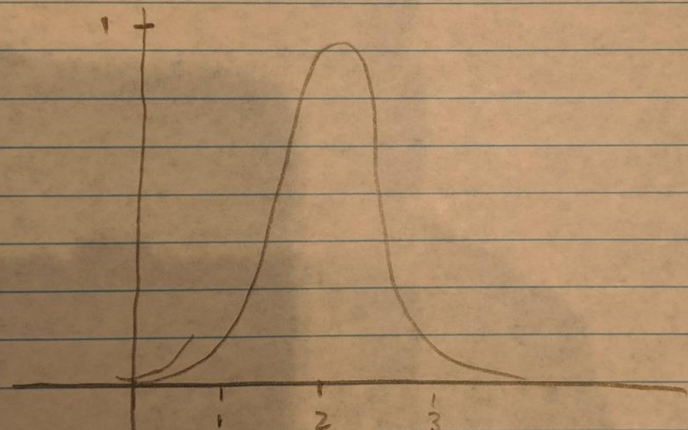
Standard Deviation (x | C2) = sqrt(1/6) = 0.4082

Q2 Plot

$C_1 \rightarrow$



$C_2 \rightarrow$



Q3

a) A classifier in this case is a function such that it takes in number x, and based on which bin it lies on, it gives the corresponding class. So a classifier can supposed to have a mapping from bin no. to a class. Considering there are k total bins (sub-intervals) the map looks something like

Bin No	Class
1	1
2	2
3	2
⋮	⋮
k	1

mapping it as a binary no. when Class 1 → 0, Class 2 → 1

a classifier → $\underset{1}{\underline{1}}\ \underset{2}{\underline{0}}\ \underset{3}{\underline{0}}\ \underset{4}{\underline{1}}\ \ldots\ \underset{k}{\underline{1}}$

So the total no. of classifiers possible in Hypothesis space H_k for given $k \rightarrow$

$\boxed{|H_k| = 2^k} \rightarrow$ total no. of binary no. possible

b) VC Dimension of H_k is equivalent $\rightarrow$ k.

Reasoning:-

Consider k points, then we can put all k points in separate bins, then in H_k there always exist a classifier for each particular configuration of these k points which need to be classified.

Consider $k+1$ points, using pigenhole principle we know that at least one bin (sub interval) would have at least two points. As soon as this happens no classifier would be able to exist or will exist in H_k such that it is able to correctly classify those two or more points if their original class is different from each other.

Hence H_k can shatter upto k points

Or formulas

$$mu1 = \frac{\sum_{i=1}^{n} X\_train(i,:) \times I\{y\_train(i) == 1\}}{\sum_{i=1}^{n} I\{y\_train(i) == 1\}}$$

$$mu2 = \frac{\sum_{i=1}^{n} X\_train(i,:) \times I\{y\_train(i) == 2\}}{\sum_{i=1}^{n} I\{y\_train(i) == 2\}}$$

$$S1 = \frac{\sum_{i=1}^{n} (X\_train(i,:) - mu1)^T \times (X\_train(i,:) - mu1) \times I\{y\_train(i) == 1\}}{\sum_{i=1}^{n} I\{y\_train(i) == 1\}}$$

$$S2 = \frac{\sum_{i=1}^{n} (X\_train(i,:) - mu2)^T (X\_train(i,:) - mu2) \times I\{y\_train(i) == 2\}}{\sum_{i=1}^{n} I\{y\_train == 2\}}$$

here X_train = Dataset of the input features from training data → Matrix

y_train = Dataset of the labels from training data, → column vector

X_train(i, :) → i^{th} row → row vector

y_train(i) → i^{th} element → one element

I{} → Identity function, n = Total number of elements in training data

m = # of elements in test data

$$\text{Error rate} = \frac{\sum_{i=1}^{n} I\{y\_test(i) \neq y\_pred(i)\}}{m}$$

```
[mu1, mu2, S1, S2, ConfuisonMatrix, ErrorRate] = classify('./data/training_data.txt',
'./data/test_data.txt')

mu1 =

    1.0554    2.5181    3.2967   -1.8927   -1.3918    4.0635   -4.3540   -5.8705

mu2 =

    3.8052    5.3740    5.7333    1.1596    1.1777    6.8000   -2.0286   -2.5044

S1 =

    0.9729    0.7135    0.4570    0.8938    0.3096    0.1975    0.7362    1.6629
    0.7135    3.0658    2.5982    0.3878    1.2994    0.1442    0.9168    4.9388
    0.4570    2.5982    6.6612    0.9084    1.6397    0.8148    0.0779    5.4168
    0.8938    0.3878    0.9084    5.0754    0.0963    1.0611    2.3978    4.5946
    0.3096    1.2994    1.6397    0.0963    2.3973   -0.0191    0.2175    2.5378
    0.1975    0.1442    0.8148    1.0611   -0.0191    1.0412   -0.0531    1.8604
    0.7362    0.9168    0.0779    2.3978    0.2175   -0.0531    6.5154    3.8609
    1.6629    4.9388    5.4168    4.5946    2.5378    1.8604    3.8609   17.0931
```

S2 =

1.3486	0.8658	0.1559	0.5847	0.9473	0.2543	0.3705	0.9435
0.8658	2.8161	-0.1819	0.2152	0.7141	0.6797	-0.2383	2.4194
0.1559	-0.1819	6.6734	1.7716	1.0318	0.6526	1.6795	3.8992
0.5847	0.2152	1.7716	3.5433	0.3570	1.3665	2.1207	3.0899
0.9473	0.7141	1.0318	0.3570	2.7517	0.1225	1.4342	2.4651
0.2543	0.6797	0.6526	1.3665	0.1225	1.8107	0.3445	1.5113
0.3705	-0.2383	1.6795	2.1207	1.4342	0.3445	7.1134	2.6246
0.9435	2.4194	3.8992	3.0899	2.4651	1.5113	2.6246	13.9151

ConfuisonMatrix =

24	11
6	59

ErrorRate =

0.1700