Homework - III

1.) a. The MLE estimate of both the coin and the thumback is the same but the MAP estimate is not. - TRUE

The MLE estimate for both the coin and the thumback depend upon the outcome of the experiment.
According to the outcome, $\alpha_H = 60$ & $\alpha_T = 40$

$$MLE_{\substack{coin \\ \& \text{ thumback}}} = \frac{\alpha_H}{\alpha_H + \alpha_T} = \frac{60}{60+40} = 0.6$$

The MAP estimates are not same due to the different beta priors.

b. The MAP estimate of the parameter θ (probability of landing heads) for the coin is greater than the MAP estimate of θ for the thumback. - FALSE

$$\text{MAP estimate for the coin/thumback} = \frac{\alpha_H + \beta_H - 1}{\alpha_H + \beta_H + \alpha_T + \beta_T - 2}$$

For coin, $\frac{60+100-1}{60+100+40+100-2} = \frac{159}{298} = 0.53356$

For thumback, $\frac{60+1-1}{60+1+40+1-2} = \frac{60}{100} = 0.6$

Thus, $MAP_{coin} < MAP_{thumback}$

2.) Let us consider a function $X \to Y$ where
$Y \to$ boolean
$X \langle X_1, X_2 \rangle$ such that X_1 is a boolean variable,
X_2 is a continuous variable.

The parameters to be estimated for each variable will be

$P(Y) \to \underline{1}$ parameter

For X_1 & Y, we need to determine

$\left.\begin{array}{l} P(X_1=0 \mid Y=0) \\ P(X_1=1 \mid Y=0) \end{array}\right\} \underline{\underline{1}}$ param ; as we can calculate the other value based on one.

$\left.\begin{array}{l} P(X_1=0 \mid Y=1) \\ P(X_1=1 \mid Y=1) \end{array}\right\} \underline{\underline{1}}$ param

$\therefore P(X_1 \mid Y) \to \underline{2}$ parameters in total.

For X_2 & Y we will need Gaussian &

$P(X_2 \mid Y=0) = \mathcal{N}(X_2 \mid \underline{\mu_{20}}\ \underline{\sigma_{20}}) \to 2$ param

$P(X_2 \mid Y=1) = \mathcal{N}(X_2 \mid \underline{\mu_{21}}, \underline{\sigma_{21}}) \to 2$ param

$\therefore P(X_2 \mid Y) \to \underline{\underline{4}}$ parameters in total

The parameters that must be estimated to define a Naive Bayes classifier in this case are $\underline{\underline{7}}$ as the formula would be

$P(Y \mid X) = (P(X \mid Y) \cdot P(Y)) / P(X) = (P(X_1 X_2 \mid Y) \cdot P(Y)) / P(X)$

$= \underline{\underline{P(X_1 \mid Y) \cdot P(X_2 \mid Y) \cdot P(Y)}}$

3.) We assume the prior probabilities for male and females as $P(M) = 0.5$ & $P(F) = 0.5$.

3) According to the given we can calculate the following mean values and variance values.

	μ_h	μ_w	μ_f	σ_h	σ_w	σ_f
Male	5.855	176.25	11.25	0.035	0.0122	0.916
Female	5.4175	132.50	7.5	0.097	0.055	1.667

$$\text{Posterior (Male)} = \frac{P(\text{Male}).P(h|\text{male}).P(w|\text{male}).P(f|\text{male})}{\text{evidence}}$$

$$\text{evidence} = P(\text{male})\,P(h|\text{male}).P(w|\text{male}).P(f|\text{male}) + P(\text{female})\,P(h|\text{female}).P(w|\text{female}).P(f|\text{female})$$

$$P(h|\text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e\left(\frac{-(6-\mu_h)^2}{2\sigma^2}\right) \approx 1.578$$

$$P(w|\text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e\left(\frac{-(130-\mu_w)^2}{2\sigma^2}\right) = 5.9881\times10^{-6}$$

$$P(f|\text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e\left(-\frac{(8-\mu_{fm})^2}{2\sigma^2}\right) = 1.3112\times10^{-3}$$

Posterior numerator (male) = 6.1984×10^{-9}

turn.

Similarly, $P(h|\text{female}) = 2.23\times10^{-1}$

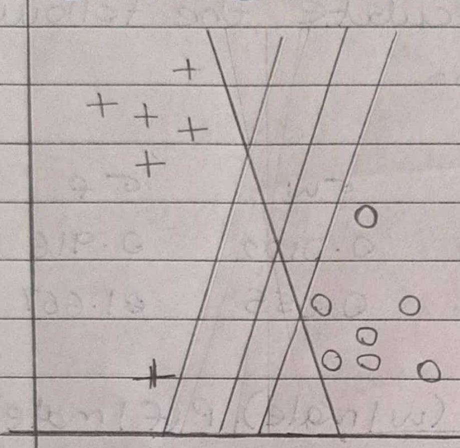
$P(w|\text{female}) = 1.67\times10^{-2}$

$P(f|\text{female}) = 2.86\times10^{-1}$

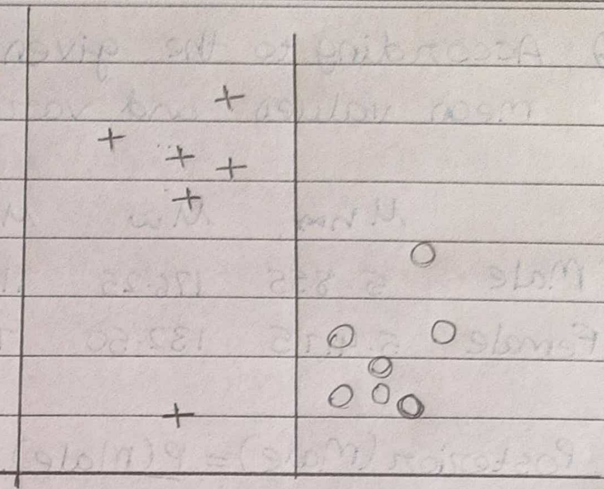
Posterior numerator (female) = 5.3778×10^{-4}

Hence, we can predict that the sample is female as num (male) is less than numerator (female).

4) a. Classification errors = 0



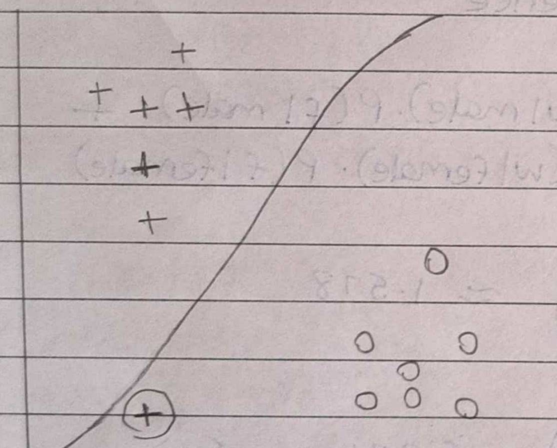
d.



Classification errors = 0

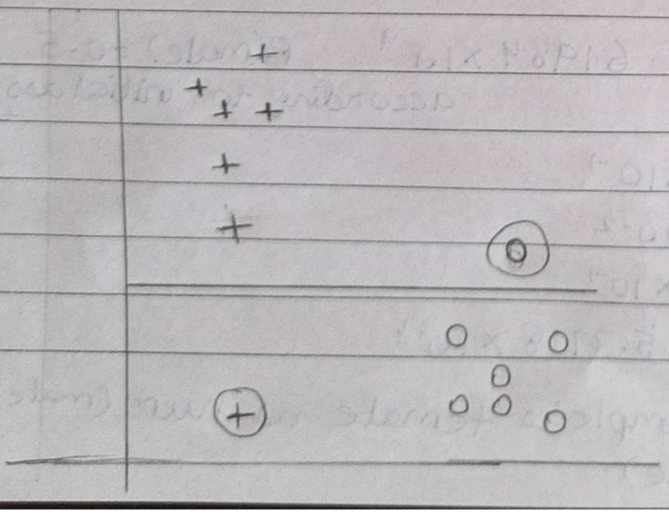
The slope of the line here is infinity, thus it is parallel to y-axis.

b.



If we regularize w_0 a lot, and don't let it grow then it is much near 0.
Classification errors = 1.

c.



Classification errors = 2.

We regularize w_1 so much, that the slope of the line is 0, thus the line is parallel to x-axis.

5.)

x_1	x_2	y
1	0	+
-1	2	-
0	-1	+

We will have three Lagragian multipliers; one for each point since there are three points.

$$L(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum \sum \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)} . x^{(j)} \rangle$$

After all the calculations, we get the dot product matrix,

	P_1	P_2	P_3
P_1	1	1	0
P_2	-1	5	-2
P_3	0	-2	1

Substituting the dot products and y-values in the equation,

$$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2(1)(1) - \frac{1}{2}\alpha_1\alpha_2(-1)(-1) - \frac{1}{2}\alpha_1\alpha_3(1)(0)$$
$$- \frac{1}{2}\alpha_1\alpha_2(-1)(-1) - \frac{1}{2}\alpha_2^2(1)(5) - \frac{1}{2}\alpha_2\alpha_3(-1)(-2)$$
$$- \frac{1}{2}\alpha_3\alpha_1(1)(0) - \frac{1}{2}\alpha_3\alpha_2(-1)(-2) - \frac{1}{2}\alpha_3^2(1)(1)$$

$$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2 - \alpha_1\alpha_2 - \frac{5\alpha_2^2}{2} - 2\alpha_2\alpha_3 - \frac{1}{2}\alpha_3^2$$

Differentiate $L(\alpha)$ w.r.t. α_1, α_2 & α_3 and set it to 0, then we get,

$$1 - \alpha_1 - \alpha_2 = 0 \quad \text{—— (1)}$$
$$1 - \alpha_1 - 5\alpha_2 - 2\alpha_3 = 0 \quad \text{—— (2)}$$
$$1 - 2\alpha_2 - \alpha_3 = 0 \quad \text{—— (3)}$$

Now, differentiate L w.r.t. w & b where

$$w = \sum_{i=1}^{m} \alpha_i y_i x_i \qquad \sum_{i=1}^{m} \alpha_i y_i = 0$$

we get $\alpha_1 - \alpha_2 + \alpha_3 = 0$ —— (4)

From eqn (4), $\alpha_2 = \alpha_1 + \alpha_3$

From eqn (2), $\alpha_2 = \dfrac{1 - \alpha_1 - 2\alpha_3}{5}$

$$\Rightarrow \quad 5(\alpha_1 + \alpha_3) = 1 - \alpha_1 - 2\alpha_3$$

$$6\alpha_1 = 1 - 7\alpha_3$$

From eqn (1) & (3),

$\alpha_2 = 1 - \alpha_1$

$\alpha_2 = \dfrac{1 - \alpha_3}{2}$ $\Rightarrow$ $2(1 - \alpha_1) = 1 - \alpha_3$

$2 - 2\alpha_1 = 1 - \alpha_3$

$1 + \alpha_3 = 2\alpha_1$

$3 + 3\alpha_3 = 6\alpha_1$

$$\therefore \quad 1 - 7\alpha_3 = 3 + 3\alpha_3$$

The values are: $\alpha_1 = 1/2$ $\quad \alpha_2 = 1/2$ $\quad \alpha_3 = 0$

Substituting values of α_1 & α_2 in w, we get

$$w = \begin{matrix} \frac{1}{2}(1)(1) + \frac{1}{2}(-1)(-1) \\ \frac{1}{2}(1)(0) + \frac{1}{2}(-1)(2) \end{matrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Substitute w in any one of the initial conditions:

$wx + b = 1$ (Positive Support vector)

$wx + b = -1$ (Negative SV)

From this substitution we get $b=0$ & $b=2$

Thus the average value of $\underline{\underline{b = 1}}$.

The α values are $\underline{\alpha_1 = 1/2}$, $\underline{\alpha_2 = 1/2}$ & $\underline{\alpha_3 = 0}$.

The weights attached to the two attributes, $\underline{w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}}$.

The bias term, $\underline{b=1}$.

6.)

x_1	x_2	class
-1	-1	-1
1	1	1
0	2	1

Let α_1, α_2 and α_3 be the lagragian multipliers for the three data points.

a. $K(x_i, x_j) = (1 + x_i^T x_j)^d$ where x_i and x_j are input vectors.

We maximize the function using

$$L(\alpha) = \sum_{i=1}^{3} \alpha_i - \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j (1 + x_i^T x_j)^2$$ such that

$\alpha_1, \alpha_2, \alpha_3 \geq 0$ & $\sum \alpha_i y_i = 0$; $-\alpha_1 + \alpha_2 + \alpha_3 = 0$.

For, $i=1, j=1$ $y_1 y_1 = (-1)(-1) = 1$

$x_1^T \cdot x_1 = [-1, -1] \begin{bmatrix} -1 \\ -1 \end{bmatrix} = 2$

$\therefore y_1 y_1 (1 + x_1^T x_1)^2 = 1(1+2)^2 = \underline{9}$.

For $i=1, j=2$, $y_1 y_2 = (-1)(1) = -1$

$$x_1^T x_2 = [-1, -1]\begin{bmatrix}1\\1\end{bmatrix} = -2$$

$$\therefore y_1 y_2 (1 + x_1^T x_2) = (-1)(1+(-2))^2 = -1$$

For $i=1, j=3$ $y_1 y_3 = (-1)(1) = -1$

$$x_1^T x_3 = [-1, -1]\begin{bmatrix}0\\2\end{bmatrix} = -2$$

$$\therefore y_1 y_3 (1 + x_1^T x_3) = (-1)(1+(-2))^2 = -1$$

Similarly, for all values of i and j $(0 < i, j \leq 3)$ we get a matrix as shown below

	P_1	P_2	P_3
P_1	9	-1	-1
P_2	-1	9	9
P_3	-1	9	25

$\Rightarrow y_i y_j (1 + x_i^T x_j)$

$$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2(9) - \frac{1}{2}\alpha_1\alpha_2 - \frac{1}{2}\alpha_1\alpha_3 - \frac{1}{2}\alpha_2\alpha_1 - \frac{1}{2}\alpha_2^2(9)$$

$$- \frac{1}{2}\alpha_2\alpha_3(9) - \frac{1}{2}\alpha_3\alpha_1 - \frac{1}{2}\alpha_3\alpha_2(9) - \frac{1}{2}\alpha_3^2(25)$$

$$\underline{\underline{L(\alpha)}} = \alpha_1 + \alpha_2 + \alpha_3 - \frac{9}{2}\alpha_1^2 - \alpha_1\alpha_2 - \alpha_1\alpha_3 - \frac{9}{2}\alpha_2^2 - 9\alpha_2\alpha_3 - \frac{25}{2}\alpha_3^2$$

b. The support vectors for the conditions $\alpha_1 = 1/8$, $\alpha_2 = 1/8$, $\alpha_3 = 0$ and $b = 0$ are the points 1 and 2 as α_1 & $\alpha_2 > 0$.

c. We Know, $w = \sum \alpha_i y^i x^i$

$$w \cdot x^t = \sum \alpha_i y^i \langle x^i . x^t \rangle = \sum \alpha_i y^i K(x^i, x^t)$$

$$= \alpha_1 y_1 K(x^1 x^t) + \alpha_2 y^2 K(x^2 x^t) + \alpha_3 y_3 K(x^3 x^t)$$

But, $\alpha_3 = 0$

$$\therefore w \cdot x^t = \alpha_1 y^1 K(x^1 x^t) + \alpha_2 y^2 K(x^2 x^t) \quad \text{--- (1)}$$

Let us calculate kernel function values,

$$K(x^i, x^j) = (1 + x_i^T x_j)^2$$

$$K(x^1, x^t) = \left(1 + [-1 \ -1]\begin{bmatrix}-1 \\ 0\end{bmatrix}\right)^2 = (1 + 1)^2 = \underline{\underline{4}}$$

$$K(x^2, x^t) = \left(1 + [1 \ 1]\begin{bmatrix}-1 \\ 0\end{bmatrix}\right)^2 = (1 - 1)^2 = \underline{\underline{0}}$$

Substituting the values in the eqn ① we get,

$$w \cdot x^t = \frac{1}{8}(-1)(4) + \frac{1}{8}(1)(0) = -\frac{1}{2} < 0.$$

We can classify the point $(x_1 = -1, x_2 = 0)$ as negative as $w \cdot x^t < 0$; it belongs class -1.

d. Let us calculate kernel function for test data point $(x_1 = 1, x_2 = 0)$

$$K(x^1, x^t) = \left(1 + [-1 \ -1]\begin{bmatrix}1 \\ 0\end{bmatrix}\right)^2 = (1 + (-1))^2 = \underline{\underline{0}}$$

$$K(x^2, x^t) = \left(1 + [1 \ 1]\begin{bmatrix}1 \\ 0\end{bmatrix}\right)^2 = (1 + 1)^2 = \underline{\underline{4}}$$

Substituting the values in the eqn ① from c, we get.

$$w \cdot x^t = \frac{1}{8}(-1)(0) + \frac{1}{8}(1)(4) = \frac{1}{2} > 0.$$

We can classify the point $(x_1 = 1, x_2 = 0)$ as positive as $w \cdot x^t > 0$; it belongs to class $+1$.