Homework 3.
Machine Learning

Name - Anushree Jumade (axj174430).

Q.1.

① The MLE estimate for both the coin and thumbtack is the same but the MAP estimate is not.

$\boxed{\text{TRUE}}$.

The MLE estimate for both depend on the outcome of the experiment. $\alpha_H = 60$ & $\alpha_T = 40$.

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

The MAP estimates are not because the beta priors are different.

② The MAP estimate of the parameter θ (probability of landing heads) for the coin is greater than the MAP estimate of θ for the thumbtack.

FALSE

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

For coin $\Rightarrow$ $\dfrac{60+100-1}{60+100+40+100-2} = \dfrac{159}{298} = 0.53356$.

$$\text{MAP estimate for the thumbtack} = \frac{60+1-1}{60+1+40+1-2} = \frac{60}{100} = 0.6.$$

Thus

$$MAP_{coin} < MAP_{thumbtack}.$$

Q.2.

a) If 'p' is the failure probability, probability of K failures before a suitable 'match' is found

$\rightarrow \quad p^k(1-p)$.

b). Log likelihood $= \sum_{i=1}^{m} \ln\left(p_i^{k} * (1-p)\right)$.

$\cancel{= \sum_{i=1}^{m} k_i \ln p_i}$

$= \sum_{i=1}^{m} \left[k_i \ln p + \ln(1-p)\right]$

$= m * \ln(1-p) + \sum_{i=1}^{m} k_i \ln(p)$ — ①

To find the p that maximizes the likelihood, we take the derivative & set to zero.

Taking derivative of ① w.r.t 'p'.

$$-\frac{m}{1-p} + \sum_{i=1}^{m}\left(\frac{k_i}{p}\right) = 0.$$

$$\sum_{i=1}^{m}\left(\frac{k_i}{p}\right) = \frac{m}{1-p}.$$

$$\frac{1}{m}\sum_{i=1}^{m} k_i = \frac{p}{1-p}$$

$$\frac{m}{\sum_{i=1}^{m} k_i} = \frac{1-p}{p}$$

$$\frac{m}{\sum_{i=1}^{m} k_i} + 1 = \frac{1-p}{p} + 1 = \frac{1}{p}$$

$$\boxed{p = \frac{\sum_{i=1}^{m} k_i}{m + \sum_{i=1}^{m} k_i}}$$

Q.3. Naive Bayes is a linear classifier

TRUE. Only under certain circumstances. When X is Gaussian & variance is class independent.

Q.4. $X \rightarrow Y$

$X < X_1, X_2 >$. $X_1 \rightarrow$ boolean $X_2 \rightarrow$ continuous variables seq.

7 parameters

2 - for Boolean X_1 -

one for $P(X_1|Y=1)$ & other for $P(X_1|Y=0)$.

4 - for Gaussian X_2. ($P(X_1|Y=1) \leftarrow 2$.

($P(X_1|Y=0)) \leftarrow 2$.

1 → for $P(Y)$.

$$P(Y|X) = \frac{P(X/Y) \cdot P(Y)}{P(X)}$$

$$= \frac{P(X_1, X_2 | Y) \cdot P(Y)}{P(X)}$$

$$= \frac{P(X_1|Y) \times P(X_2|Y) * P(Y)}{P(X)}.$$

Q.5

F1	F2	F3	Category
a	c	a	+
c	a	c	+
a	a	c	−
b	c	a	−
c	c	b	−

$P(\text{Class }+) = 2/5 \quad P(\text{Class }-) = 3/5.$

$P(F1 = a \mid \text{Class }+) = 1/2 \quad P(F1 = a \mid \text{Class }-) = 1/3$

$P(F1 = b \mid \text{Class }+) = 0 \quad P(F1 = b \mid \text{Class }-) = 1/3$

$P(F1 = c \mid \text{Class }+) = 1/2 \quad P(F1 = c \mid \text{Class }-) = 1/3$

$P(F2 = a \mid \text{Class }+) = 1/2 \quad P(F2 = a \mid \text{Class }= -) = 1/3$

$P(F2 = b \mid \text{Class }+) = 0 \quad P(F2 = b \mid \text{Class }-) = 0$

$P(F2 = c \mid \text{Class }+) = 1/2 \quad P(F2 = c \mid \text{Class }-) = 2/3$

$P(F3 = a \mid \text{Class }+) = \cancel{0}\,1/2 \quad P(F3 = a \mid \text{Class }-) = 1/3$

$P(F3 = b \mid \text{Class }+) = 0 \quad P(F3 = b \mid \text{Class }-) = 1/3$

$P(F3 = c \mid \text{Class }+) = 1/2 \quad P(F3 = c \mid \text{Class }-) = 1/3.$

$F1 = a \quad F2 = c \quad F3 = b$

$P(F1 = a \mid \text{Class }+) * P(F2 = c \mid \text{Class }+) * P(F3 = b \mid \text{Class }+) * P(\text{Class }+)$

$= \frac{1}{2} \times \frac{1}{2} \times 0 \times \frac{2}{5} = 0.$

$P(F1 = a \mid \text{Class }-) * P(F2 = c \mid \text{Class }-) * P(F3 = b \mid \text{Class }-) * P(\text{Class }-)$

$= \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{3}{5} = \frac{2}{45}$

Therefore, Naive Bayes would classify the given case in (Class −).

6. Naive Bayes.

Person	height	weight	foot size
male	6	180	12
male	5.92	190	11
male	5.58	170	12
female	~~5.92~~ 5	100	6
female	5.5	150	8
female	5.42	150	7
female	5.75	130	9
male	5.92	165	10

To Classify → $\boxed{6,\ 130,\ 8}$

Solⁿ $P(M) = 0.5$ & $P(F) = 0.5$. This prior probability could be based on our knowledge of freq. in the larger population or on frequency in the training set.

Stats derived from given training data:

	Mean Height	Mean Wt	Mean ft size	Variance ht	Variance wt	Variance ft
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

$$\text{Posterior (Male)} = \frac{P(\text{male})\, P(ht \mid \text{male})\, p(wt \mid \text{male})\, p(ft\ size \mid \text{male})}{\text{evidence}}$$

evidence = P(male) p(ht | male) · p(wt | male) · p(ft sz | male)
+ P(female) p(ht | female) p(wt | female) p(ft sz | female))

$p(ht|male) = \frac{1}{\sqrt{2\pi\sigma^2}} e\left(\frac{-(6-\mu_m)^2}{2\sigma^2}\right) \approx 1.5789$ ← probability density because ht is a continuous variable

$p(wt|male) = \frac{1}{\sqrt{2\pi\sigma^2}} e\left(\frac{-(130-\mu_m)^2}{2\sigma^2}\right) = 5.9881 \times 10^{-6}$

$P(\text{ft size}|male) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-}\left(\frac{-(8-\mu)^2}{2\sigma^2}\right) = 1.3112 \times 10^{-3}$

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

P(female) = 0.5

p(ht | female) = 2.23×10^{-1}

P(wt | female) = 1.67×10^{-2}

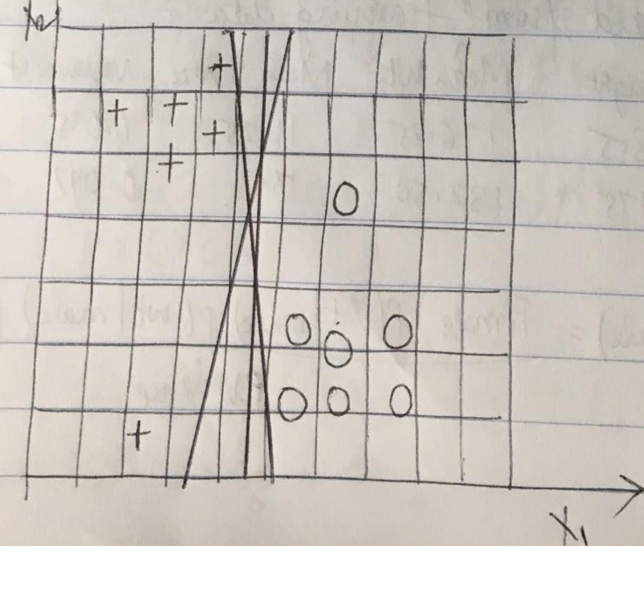
p(ft size | female) = 2.86×10^{-1}

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

Since numerator (male) is less than the one in female case, we predict sample is **FEMALE**

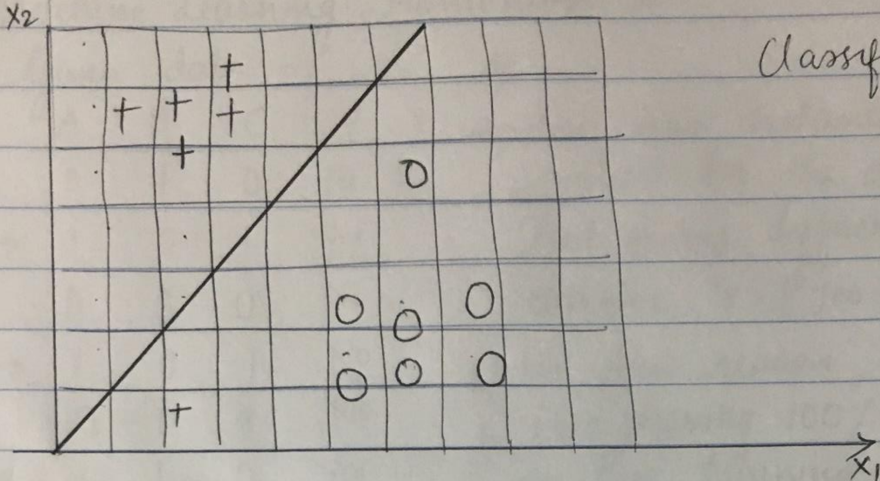
Q.7.

a).



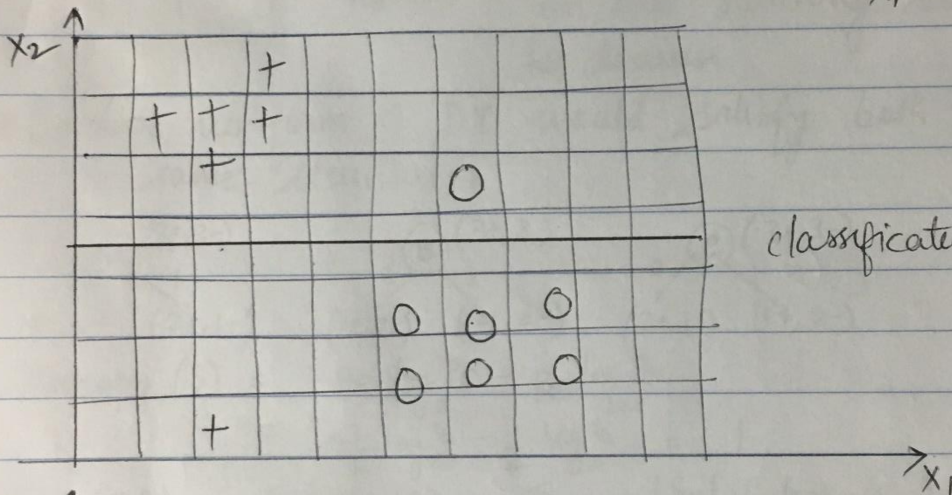
Classification errors = 0

(b)



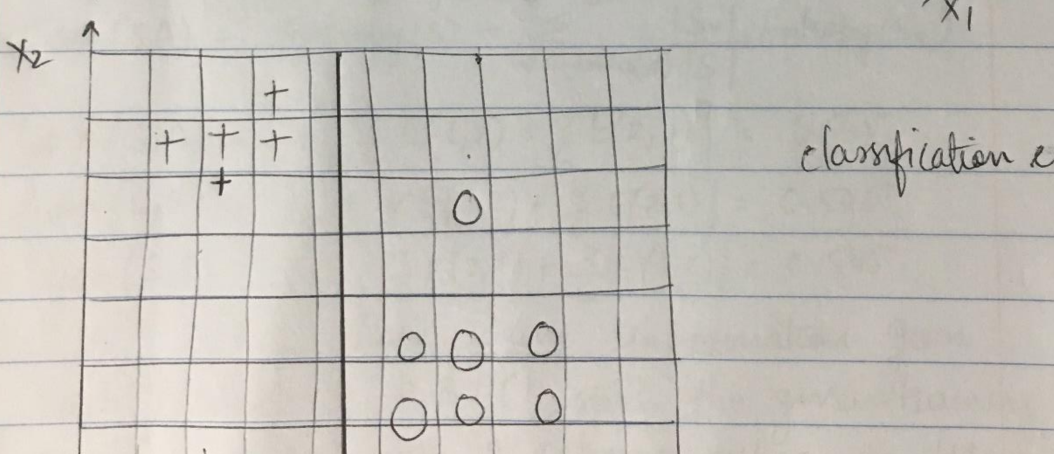
Classification errors = 1.

(c)



classification errors = 2.

(d)



classification error(s) = 0