ANEEK DAS

CSC 2515 - Homework 1

Q1) a) 2 independent, univariate R.V.'s X & Y sampled uniformly from $[0,1]$.

Find $E(Z)$ and $Var(Z)$, where

$$Z = (X-Y)^2$$

Since, X & Y are sampled from a uniform dist.

$$E(X) = \int_0^1 x.f(x)\,dx \quad ; \quad f(x) = \frac{1}{b-a} = \frac{1}{1-0} = 1$$

$$= \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

$$E(Y) = \int_0^1 y\,f(y)\,dy \quad ; \quad f(y) = \frac{1}{1-0} = 1$$

$$= \left[\frac{y^2}{2}\right]_0^1 = \frac{1}{2}$$

$$E(X^2) = \int_0^1 x^2 f(x)\,dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

$$E(Y^2) = \int_0^1 y^2 f(y)\,dy = \left[\frac{y^3}{3}\right]_0^1 = \frac{1}{3}$$

Now, $E(Z) = E\left[(X-Y)^2\right] = E(X^2) + E(Y^2) - 2E(XY)$

Since, X & Y are independent

$$E(XY) = E(X).E(Y)$$

putting the values.

$$E\left[(X-Y)^2\right] = \frac{1}{3} + \frac{1}{3} - \cancel{2}\left(\frac{1}{\cancel{2}}\right).\left(\frac{1}{2}\right)$$

$$= \frac{2}{3} - \frac{1}{2} = \frac{4-3}{6} = \frac{1}{6}$$

So, $E\left[(X-Y)^2\right] = \frac{1}{6}$ ← Ans.

Now, to find Variance.

$$Var(Z) = Var[(X-Y)^2]$$

$$= Var(X^2 + Y^2 - 2XY)$$

$$= Var(X^2) + Var(Y^2) + 4Var(XY) + 2cov(X^2, Y^2) - 4cov(X^2, XY) - 4cov(Y^2, XY)$$

Now, since, X & Y are independent.

$$Var(Z) = Var(X^2) + Var(Y^2) + 4Var(XY) \quad \text{—— ①}$$

Now, we need to find the individual terms in ①.

$$Var(X^2) = E(X^{2^2}) - [E(X^2)]^2$$

$$= \underbrace{E(X^4)} - [E(X^2)]^2$$

$$\int_0^1 x^4 f(x)\,dx = \left[\frac{x^5}{5}\right]_0^1 = \frac{1}{5}.$$

So, $Var(X^2) = \frac{1}{5} - \left(\frac{1}{3}\right)^2$

$$= \frac{1}{5} - \frac{1}{9} = \frac{9-5}{45} = \frac{4}{45}$$

$$Var(Y^2) = \underbrace{E(Y^4)} - [E(Y^2)]^2$$

$$\int_0^1 y^4 f(y)\,dy = \left[\frac{y^5}{5}\right]_0^1 = \frac{1}{5}$$

$$Var(Y^2) = \frac{1}{5} - \left(\frac{1}{3}\right)^2 = \frac{4}{45}$$

Now, $Var(X.Y) = E(X^2)E(Y^2) - (E(X))^2(E(Y))^2$

$$= \left(\frac{1}{3} \times \frac{1}{3}\right) - \left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$$

$$= \frac{1}{9} - \frac{1}{16} = \frac{7}{144}$$

putting the values in ① we get.

$$Var(Z) = \frac{4}{45} + \frac{4}{45} + \cancel{4} \times \frac{7}{\cancel{144}_{36}} = \frac{8}{45} + \frac{7}{36} = \frac{67}{180} \leftarrow \text{Ans}$$

Q1) b). $R = Z_1 + Z_2 + \dots + Z_d$, where $Z_i = (X_i - Y_i)^2$.

find $E[R]$ and $Var[R]$

$$E[R] = E[Z_1 + Z_2 + \dots + Z_d]$$

$$= E[Z_1] + E[Z_2] + \dots + E[Z_d].$$

Now, since all the pts. are uniformly distributed b/w [0,1] in the 'd' dimensions

$$E[z_1] = E[z_2] = \dots = E[z_d].$$

So, $E[R] = d \times E[z_1]$

Now, from Q1)a) we know $E[z] = 1/6$.

So, $E[R] = d \times \frac{1}{6} = \underline{\underline{\frac{d}{6}}} \xleftarrow{\text{Ans}}$

Now, for $var(R)$

$$= var[z_1 + z_2 + \dots + z_d]$$

Now, since, $z_1, z_2, \dots, z_d$ are pairwise uncorrelated.

$$var[z_1 + z_2 + \dots + z_d] = var[z_1] + var[z_2] + \dots + var[z_d].$$

Now, from Q1)a) we know $var[z] = \frac{67}{180}$. and all the variances are equal

So, $var(R) = d \times var[z_1]$

$$= \underline{\underline{\frac{67d}{180}}} \xleftarrow{\text{Ans}}$$

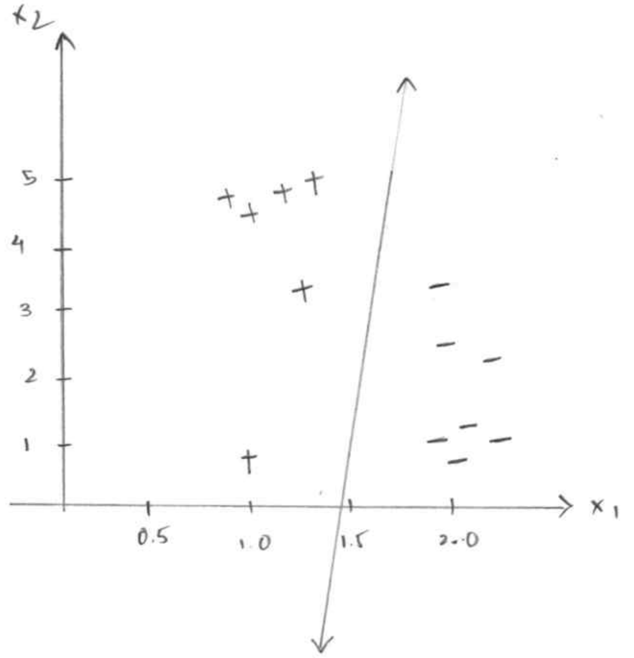
Q2> a> $P(y=1 \mid x, w) = \sigma(w_0 + w_1 x_1 + w_2 x_2)$

Fit model by maximum likelihood i.e, minimizing the cross entropy loss.

$$J(w) = -\log Pr(D_{train}; w)$$

where $D_{train} = \{(x^{(i)}, y^{(i)})\}$ i = 1 to n for n points in the training set.

Sketch a possible decision boundary.



Q2> b> The decision boundary is not unique.

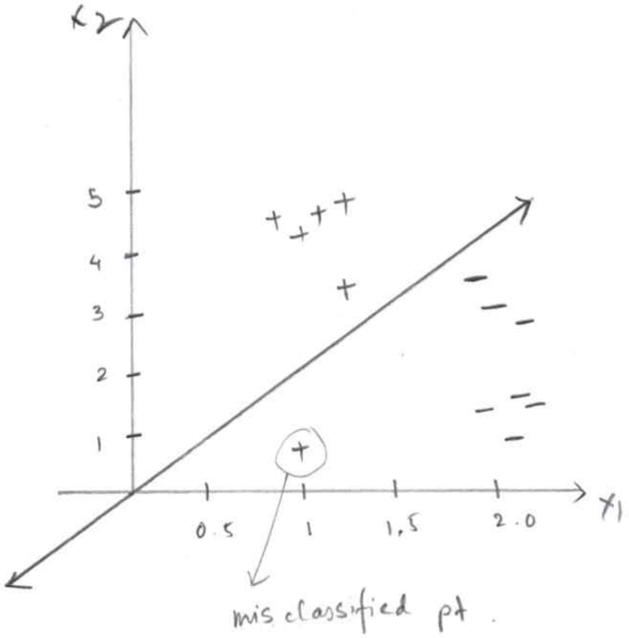
Q2> c> 0 classification errors on the training set.

Q2> d> Regularize the w_0 parameter, so, we minimize

$$J(w) = -\log P(D_{train}; w) + \lambda w_0^2$$

Now, when $\lambda \to \infty$, $w_0 \to 0$ as even a small value of w_0 will result in a huge loss.

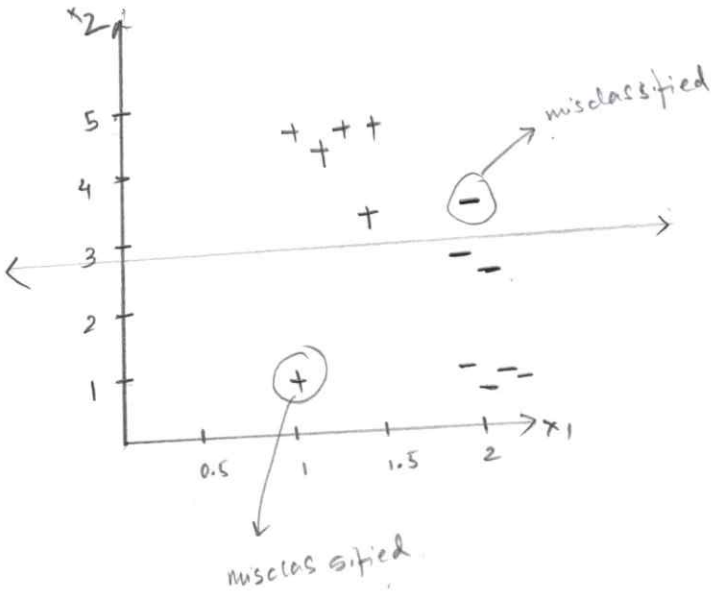
Q2> e> One classification error on the training set.



Q2> f> Regularize the 'w_1' parameter; we minimize

$$J(w) = -\log P(D_{train} ; w) + \lambda w_1^2$$

As $\lambda \to \infty$, $w_1 \to 0$. So, we can only use w_0 & w_2.

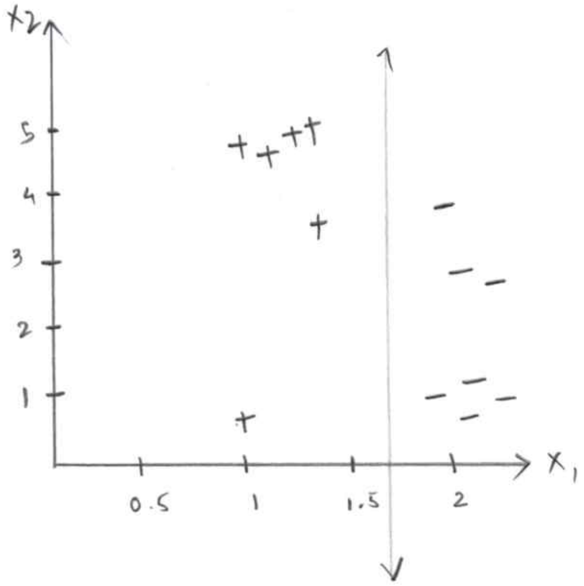


Q2> g> 2 classification errors on the training set.

Q2> h> Regularize the 'w_2' parameter; we minimize.

$$J(w) = -\log P(D_{train} ; w) + \lambda w_1^2$$

As $\lambda \to \infty$, $w_2 \to 0$.



Q2> i> 0 classification errors on the training set.

Q3) a) Prove that entropy $H(x)$ is non-negative.

For a discrete random variable X with P.M.F 'p'

$$H(X) = \sum_x p(x) . \log_2 \left(\frac{1}{p(x)} \right)$$

where, $p(x)$ is the prob. of $X = x$. Now, $\frac{1}{p(x)}$ is the factor of uncertainty reduction.

$\log_2 \left(\frac{1}{p(x)} \right)$ gives us the number of bits of useful information conveyed, when we reduce the uncertainty.

We know $0 \leq p(x) \leq 1$

So, $$H(x) = - \sum_x p(x) . \underbrace{\log_2 (p(x))}$$.

Now, log (fraction) is negative.
and $p(x)$ is positive.

So, $$H(x) = - \sum (\text{+ve value}) (\text{-ve value}).$$
$$= - \sum (\text{-ve values})$$
$$= \text{+ve}.$$

So, $H(x)$ is non-negative.

Q3) b) Prove that $KL(p \| q)$ is non-negative.

$p(x) \longrightarrow$ true prob. of distribution

$q(x) \rightarrow$ we are trying to fit 'q' on distribution 'p'.

The KL divergence measures the difference / divergence when we try to fit 'q' on 'p'.

$$KL(p \| q) = \sum_x p(x) . \log_2 \frac{p(x)}{q(x)}.$$

$$\Rightarrow -KL(p \| q) = - \sum_x p(x) . \log \frac{p(x)}{q(x)}.$$

$$\Rightarrow \quad '' \quad = \sum_x p(x) \log \frac{q(x)}{p(x)} \quad \text{--- (1)}$$

Now, Jensens Inequality states.

$$f(E[X]) \leq E[f(X)] \text{ for convex function } f(x).$$

Now, log is a strictly concave function. So,

$$E[f(x)] \leq f(E[x]).$$

So, pulling the log out we get.

$$-KL(p\|q) \leq \log \sum_x p(x) \cdot \frac{q(x)}{p(x)}$$

$$\leq \log \sum_x q(x)$$ [only possible when the set of x includes all possible values of $X = x$]

for all values of x

$$\leq \log \sum_{x \in X} q(x).$$

Now, $\sum_{x \in X} q(x) = 1$ as it includes the summation of probabilities of all possible events

$$= \log 1 = 0.$$

If both the distributions 'p' and 'q' are same. i.e, $p(x) = q(x)\ \forall x$ then the divergence will be 0. Else, it will be +ve.

Q3) c) Information Gained / Mutual Info. b/w X & Y:

$$I(Y; X) = H(Y) - H(Y|X)$$

Prove:

$$I(Y; X) = KL(p(x,y) \| p(x) \cdot p(y))$$

Mutual Info. gives us the error of using $p(x)\,p(y)$ to model the joint prob. dist $p(x,y)$ when x & y are independent of each other.

$$I(Y;X) = H(Y) - H(Y|X)$$

$$-\sum_y p(y) \log p(y) \quad - \quad -\sum p(x,y) \log p(y|x)$$

$$I(Y;X) = -\sum_{y} p(y)\log p(y) - \left(-\sum p(x,y)\log p(y|x)\right) \quad \text{— (1)}$$

Now, $p(x) = \sum_{y} p(x,y) \rightarrow$ marginal dist. of X. (Given).

Similarly we can write

$$p(y) = \sum_{x} p(x,y) \rightarrow \text{marginal dist of } Y. \quad \text{— (2)}$$

using (2) in (1) we get $\rightarrow$

$$I(Y;X) = -\sum_{x,y} p(x,y)\log p(y) + \sum_{x,y} p(x,y)\log p(y|x).$$

$$= -\sum_{x,y} p(x,y)\log \frac{p(y|x)}{p(y)}. \quad \text{— (3)}$$

Now, $p(y|x) = \dfrac{p(x,y)}{p(x)}$. — (4).

Putting (4) in (3)

$$= -\sum_{x,y} p(x,y)\log \frac{p(x,y)}{p(x)\cdot p(y)}.$$

$$= KL(p(x,y)\,\|\,p(x)\cdot p(y)).$$