Name: Venkata Subba Narasa Bharath, Meadam

SBU ID: 112672986

CSE-512 (HW 4)

1)

(a) $P(x = red) = \frac{1}{2}$ $\quad$ $P(x = yellow) = \frac{1}{5}$

$P(x = blue) = \frac{1}{4}$ $\quad$ $P(x = black) = \frac{1}{20}$

$$\boxed{H(x) = -\sum_{x=x} Pr(x=x) \log_2 (Pr(x=x))}$$

$$= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{1}{5}\log_2\left(\frac{1}{5}\right) + \frac{1}{20}\log_2\left(\frac{1}{20}\right)\right)$$

$= 1.68$ bits

b) X = color of a sock randomly picked

Y = which drawer (up, or down)

$P(\text{top drawer}) = 2\,P(\text{bottom drawer})$ — (1)

$P(\text{top drawer}) + P(\text{bottom drawer}) = 1$ — (2)

Solving the above 2 equations, we get

$$\boxed{\begin{aligned} &P(\text{top drawer}) = 2/3 \\ &P(\text{bottom drawer}) = 1/3 \end{aligned}}$$

$$\boxed{H(X|Y) = -\sum_{X=x, Y=y} Pr(X=x, Y=y) \log_2 (Pr(X=x|Y=y))}$$

calculating relevant Probabilities

$Pr(X=x, Y=y) = Pr(X|Y) * Pr(Y)$

$Pr(X=red | Y=top) = 1$

$Pr(x=red, y=top) = 1 \times \frac{2}{3} = \frac{2}{3}$

$Pr(X=red | Y=bottom) = 0$

$Pr(X=blue | Y=top) = 0$

$Pr(x=blue | Y=below) = \frac{1}{2}$

$Pr(x=blue, Y=below) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$

$Pr(x = \text{yellow} \mid y = \text{top}) = 0$

$Pr(x = \text{yellow} \mid y = \text{below}) = 2/5$

$Pr(x = \text{yellow}, y = \text{below}) = \frac{2}{5} \times \frac{1}{3} = \frac{2}{15}$

$Pr(x = \text{black} \mid y = \text{top}) = 0$

$Pr(x = \text{black} \mid y = \text{below}) = \frac{1}{10}$

$Pr(x = \text{black}, y = \text{below}) = \frac{1}{10} \times \frac{1}{3} = \frac{1}{30}$

we, will assume, $0 \log_2(0)$ as 0

$$H(X|Y) = -\left[\frac{2}{3}\log_2(1) + \frac{1}{6}\log_2\left(\frac{1}{2}\right) + \frac{2}{15}\log_2\left(\frac{2}{5}\right) + \frac{1}{30}\log_2\left(\frac{1}{10}\right)\right]$$

$$= 0.45 \text{ bits}$$

c) Information gain:-

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

$$= 1.68 - 0.45$$

$$= 1.23$$

2)

a) $p(\text{word} = \text{the}) = 6/141$

$p(\text{word} = \text{rabbit}) = 3/141$

$p(\text{word} = a) = 5/141$

i) $Pr(\text{current word} = \text{rabbit} \mid \text{Previous word} = \text{the})$

$\Rightarrow \frac{\cancel{2}^{1}}{\cancel{6}_{3}} = \frac{1}{3}$

ii) $Pr(\text{current word} = a \mid \text{Previous word} = \text{rabbit})$

$= \frac{1}{3}$

iii) $Pr(\text{current word} = \text{the} \mid \text{Previous word} = \text{rabbit})$

$= \frac{2}{3}$

iv) $Pr(\text{current word} = \text{the or a} \mid \text{previous word} = \text{rabbit})$

$= 1$

v) $P(A \cap B) = P(A) * P(B) \rightarrow$ condition for naive Baye's

$p(\text{current word} = \text{the} \cap \text{previous\_word} = \text{rabbit})$

$= \frac{2}{141}$

$p(\text{word} = \text{the}) = \frac{6}{141}$

$p(\text{word} = \text{rabbit}) = \frac{3}{141}$

$\Rightarrow \frac{2}{141} \neq \frac{6}{141} \times \frac{3}{141}$

$\boxed{\therefore \text{ Naive Bayes assumption is not valid here}}$

Challenge

Q3) $E[x] = \sum_{i=1}^{n} x_i p_i$

$E[f(x)] = \sum_{i=1}^{n} p_i f(x_i)$

$f(E[x]) = f(x_1 p_1 + x_2 p_2 \dots x_n p_n)$

$f(E(x)) = f\left(p_1 x_1 + (1-p_1)\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]\right)$

$\leq p_1 f(x_1) + (1-p_1) f\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]$

(Assuming $\theta = p_1$)

$f(E[x]) \leq p_1 f(x_1) + (1-p_1) f\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]$ — (1)

Let us consider the term,

$= f\left[\frac{p_2}{1-p_1} x_2 + \frac{p_3}{1-p_1} x_3 \dots \frac{p_n}{1-p_n} x_n\right]$

$= f\left(\frac{p_2 x_2}{1-p_1} + \frac{1-p_1-p_2}{1-p_1}\left[\frac{p_3 x_3}{1-p_1-p_2} + \frac{p_4 x_4}{1-p_1-p_2} \dots + \frac{p_n x_n}{1-p_1-p_2}\right]\right)$

$$< \frac{P_2}{1-P_1} f(x_2) + \frac{1-P_1-P_2}{1-P_1} f\left[\frac{P_3x_3}{1-P_1-P_2} + \frac{P_4x_4}{1-P_1-P_2} \cdots + \frac{P_nx_n}{1-P_1-P_2}\right]$$

$$\theta = \frac{P_2}{1-P_1} \qquad 1-\theta = 1-\frac{P_2}{1-P_1} = \frac{1-P_1-P_2}{1-P_1}$$

$$\Rightarrow (1-P_1) f\left(\frac{P_2x_2}{1-P_1} + \frac{P_3x_3}{1-P_1} \cdots \frac{P_nx_n}{1-P_1}\right)$$

$$< P_2 f(x_2) + (1-P_1-P_2) f\left(\frac{P_3x_3}{1-P_1-P_2} \cdots + \frac{P_nx_n}{1-P_1-P_2}\right) \quad \text{— (2)}$$

From (1) & (2), we get,

$$f(E[x]) < P_1 f(x_1) + P_2 f(x_2) + (1-P_1-P_2) f\left(\frac{P_3x_3}{1-P_1-P_2} \cdots + \frac{P_nx_n}{1-P_1-P_2}\right)$$

If we generalize the above term,

we get

$$f(E[x]) < P_1 f(x_1) + P_2 f(x_2) + P_3 f(x_3) \cdots + P_{n-1} f(x_{n-1}) + (1-P_1-P_2\cdots-P_{n-1}) f\left(\frac{P_nx_n}{1-P_1-P_2\cdots-P_{n-1}}\right)$$

$\Rightarrow p_1 + p_2 + p_3 \ldots \quad p_n = 1$

$\Rightarrow 1 - p_1 - p_2 - p_3 \ldots - p_{n-1} = p_n$

$f(E[x]) \leq p_1 f(x_1) + p_2 f(x_2) \ldots$

$+ p_{n-1} f(x_{n-1}) + p_n f(x_n)$

$$\boxed{\therefore f(E[x]) \leq E[f(x)]}$$

Q4) $P_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$

(a) Likelihood (L),

$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$

$= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$

$\log(L(\lambda)) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$

For maximum, set derivative to '0'.

$$\frac{dL(\lambda)}{d\lambda} = \frac{n}{\lambda} - \Sigma x_i = 0$$

$$\boxed{\begin{gathered}\lambda_{MLE} = \frac{n}{\Sigma x_i} \\ \lambda_{MLE} = \frac{1}{\bar{\theta}} \quad \left(\bar{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}\right)\end{gathered}}$$

b) $E[\hat{\lambda}] = E\left[\frac{n}{\sum_{i=1}^{n} x_i}\right]$

$$= n \times E\left[\frac{1}{x_i}\right]$$

$$E[x^{-1}] = \int_0^\infty \frac{1}{\lambda} \lambda e^{-\lambda x} dx = \frac{\lambda}{n-1}$$

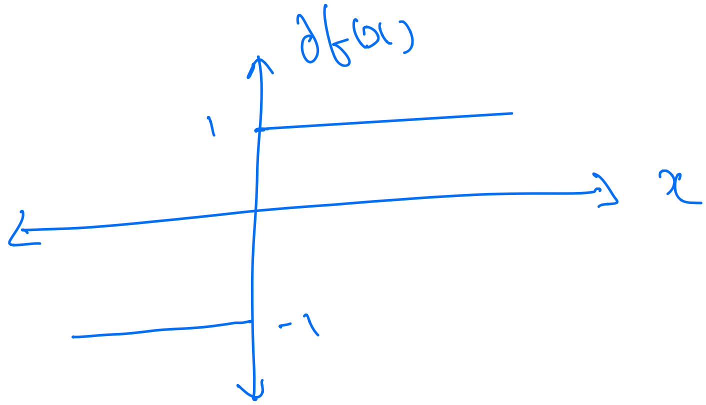
$$\boxed{\begin{gathered}E(\hat{\lambda}) = \frac{n}{n-1} \lambda \\ \neq \lambda\end{gathered}}$$

$\rightarrow$ It is a biased estimator

Q2)

(a) If f is convex and diffrentiable, then gradient at x is a sub gradient.

But a sub-gradient can exist even when f is non-diffrentiable.



A function f is called sub-diffrentiable at x if there exists atleast one sub-gradient at x.

Consider, $f(x) = |x|$

For, $x < 0 \rightarrow$ sub-gradiet $\partial f(x) = -1$

For, $x > 0 \longrightarrow$ sub-gradiet $\partial f(x) = 1$

At $\boxed{x \; = 0}$

one sub-gradient is defined by the equality, $\boxed{|z| \geq gz}$ $\forall z$ which is satisfied iff $\boxed{g \in [-1, 1]}$.

$\therefore \partial f(0) = [-1, 1]$

$$\therefore \partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \{-1\} & \text{if } x < 0 \\ [-1, 1] & x = 0 \end{cases}$$

b) A point x^* is a minimizer of a convex function iff f is sub-differentiable at x^* and $0 \in \partial f(x^*)$,

i.e. $g = 0$ is a sub-gradient of f at x^*.

[as $f(x) \geq f(x^*)$]

$\Rightarrow 0 \in \partial f(x^*)$ reduces to $\nabla f(x^*) = 0$

if f is differentiable at x^*.