Advanced Machine Learning

HW: 1

(1 LATE DAY)

See A

Problem 1:

Q1. Since the dis^n ~~is~~ an i.i.d and sampled from Bernoulli.
Let the likelihood f^n be:

$$Li(Ber(p)) = \prod_{i=1}^{n} p^{(x_i)}(1-p)^{(1-x_i)}$$

$\therefore$ MLE is

$$\frac{\partial Li}{\partial p} = 0$$

Taking log: $\log(Li) = \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i)$

$\because$ it is a log ~~...~~ f^n $MLE(Li) = MLE(\log(Li))$

$\therefore$ $\frac{\partial(\log(Li))}{\partial p} = 0$

$$\Rightarrow \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(1-x_i)}{(1-p)} = 0$$

$$\Rightarrow (1-p)\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n}(1-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - np = 0$$

$$\Rightarrow \boxed{p = \frac{\sum_{i=1}^{n} x_i}{n}}$$

$\therefore$ To further prove that this is the max. & not min. value

$$\frac{\partial^2(\log(Li))}{\partial p^2} < 0 \rightarrow \text{this should hold}$$

$$= -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{\sum_{i=1}^{n}(1-x_i)}{(1-p)^2}$$

$$= \frac{-\sum_{i=1}^{n} x_i}{p^2} - \frac{(n - \sum_{i=1}^{n} x_i)}{(1-p)^2}$$

$p \in [0,1] \rightarrow$ prob. value

& $x_i \in \{0,1\} \rightarrow$ observ^n from Bernoulli dis^n

$$\therefore \frac{-\sum_{i=1}^{n} x_i}{p^2} - \frac{(n - \sum x_i)}{(1-p)^2} < 0$$

$\therefore$ $p = \frac{\sum_{i=1}^{n} x_i}{n}$ is indeed MLE of Ber(p)

Q.2)

1) $$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Let the likelihood f^n be : $Li(\lambda; x_1 \ldots x_n)$

$$\therefore Li(\lambda; x_1 \ldots x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Taking ~~log~~ ln

$$\ln(Li) = \sum_{i=1}^{n} \left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \right)$$

$$= -\ln(x_i!) + \ln(\lambda) \sum_{i=1}^{n} x_i - n\lambda$$

$$\text{MLE} \Rightarrow \frac{\partial(Li)}{\partial \lambda} = \frac{\partial(\ln(Li))}{\partial \lambda} = 0$$

$$\Rightarrow -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

$$\boxed{\Rightarrow \lambda = \frac{\sum_{i=1}^{n} x_i}{n}}$$

To further prove, it is indeed max value of λ

$$\frac{\partial^2 Li}{\partial \lambda^2} < 0 \rightarrow \text{must hold}$$

$$\therefore \frac{\partial^2 \ln(Li)}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$$

$x_i \in [0, 1] \rightarrow$ poisson ds^n

& $\lambda > 0$ for a poisson ds^n

$$\therefore \frac{\partial^2 \ln(Li)}{\partial \lambda^2} < 0$$

Thus, $\boxed{\lambda = \frac{\sum_{i=1}^{n} x_i}{n}}$ is the MLE for $P(x)$.

b) Expectation value for a discrete random variable like Y is :

$$E(Y) = \sum_{\text{for all } y} y \, Pr(Y=y)$$

$Pr(Y=y)$ is the probability dsn fn

∴ for Poisson dsn

$$E(Y) = \sum_{y=0}^{n} \frac{y \lambda^y e^{-\lambda}}{y!}$$

$$\rightarrow E(y) = \lambda e^{-\lambda} \sum_{i=1}^{n} \frac{\lambda^{i-1}}{(i-1)!}$$

∵ at $y=0$, the term becomes 0

$$\Rightarrow E(y) = \lambda e^{-\lambda} \sum_{x=0}^{n-1} \frac{\lambda^x}{x!} \quad (\text{substituting } x = i-1)$$

$$\Rightarrow E(y) = \lambda e^{-\lambda} \left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots \right]$$

∵ it is a Taylor series it's expansion is given by

$$f(0) + \frac{f'(0)x}{1!} + \frac{f''(0)x^2}{2!} + \cdots$$

Thus, we get

$$E(y) = \lambda e^{-\lambda} \left[1 + \lambda + \frac{\lambda^2}{2!} + \cdots \right]$$

Using the previous Taylor expansion

$E(y) = \lambda e^{-\lambda}(e^{\lambda})$

$\boxed{\Rightarrow E(y) = \lambda}$ → expectation of Poisson dsn

Problem : 2

Q1) The mean & std. deviation of $x_1, x_2 \ldots x_n$ samples is given as :

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N} \quad \& \quad \sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{(N-1)}}$$

(mean) (std. dev)

For "no of wheels" feature :

$\mu_1 = \frac{4+4+2+8+4+3}{6} = 4.1667$

$\sigma_1 = \sqrt{\frac{(4-4.1667)^2 + \ldots (3-4.1667)^2}{5}} = 2.04$

Normalizing the data points gives us :

$-0.082, -0.082, -1.062, 1.878, -0.082$ and -0.572

Similarly "cost dollars" feature.

$\mu_2 = 20666.67$ $\sigma_2 = 11707.55$

After normalizing this feature:

$-0.484, 0.37, -1.338, 1.651, 0.114, -0.313$

Q2)

$$J(w_0, w_1, w_2) = \frac{1}{m} \sum_{i=1}^{m} \left(h_w(x^{(i)}) - y^{(i)}\right)^2$$

$\rightarrow$ $h_w(x^{(i)}) = w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)}$

$m = 6$

$x_1 =$ "no of wheels"

$x_2 =$ " cost (dollars) "

Q3. In order to minimise the error between $h_w(x)$ and ground truth y, which is our aim, least squares is useful. This is because taking the square deviation penalizes the ~~both~~ positive & negative deviations equally. Furthermore, squaring penalizes higher errors more.

Q4.

$$J(w) = \frac{1}{6} \sum_{i=1}^{6} \left(h_w(x)^{(i)} - y^{(i)} \right)^2$$

$$= \frac{1}{6} \sum_{i=1}^{6} \left[\left(w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)} \right) - y^{(i)} \right]^2$$

$$\therefore \frac{\partial J}{\partial w_0} = \frac{2}{6} \sum_{i=1}^{6} \left[\left(w_0 + w_1 x_1^{(i)} + w_2 x_2^{(i)} \right) - y^{(i)} \right]$$

$$= \frac{1}{3} \sum_{i=1}^{6} \left[h_w(x)^{(i)} - y^{(i)} \right]$$

$$\frac{\partial J}{\partial w_1} = \frac{2}{6} \sum_{i=1}^{6} x_1^{(i)} \left[h_w(x)^{(i)} - y^{(i)} \right]$$

$$= \frac{1}{3} \sum_{i=1}^{6} x_1^{(i)} \left[h_w(x)^{(i)} - y^{(i)} \right]$$

$$\frac{\partial J}{\partial w_2} = \frac{1}{3} \sum_{i=2}^{6} x_2^{(i)} \left[h_w(x)^{(i)} - y^{(i)} \right]$$

Q5) Model (i) →

$$h_w(x) = w_0$$

This model has a high bias & low variance. Since it is just a constant value it will underfit and will lead to high train & test error.

Model (2) →

$$h_w(x) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5$$

This model has low bias and high variance as we rely too much on a single feature. Since it is a complex model it will eventually lead to overfitting. Thus, we will get low training but high test error.

Q6) Model (1) performance would improve with some dependance on features, some sort of linear or polynomial relation with features. While model (2) can be improved by reducing its complexity to maybe second or third degree and introducing an additional feature.

Problem 3:

Q1) We are given:

$C_0 \rightarrow$ infectious
$C_1 \rightarrow$ non-infectious
$X \rightarrow$ symptoms

Thus, $P(C_0|X) \rightarrow$ Probability that the disease is infectious given the symptoms

$P(X|C_0) \rightarrow$ Probability that one has certain symptoms when they have an infectious disease

$P(C_0) \rightarrow$ Probability that disease is infectious

Q2) Since, there are two classes C_0 & C_1

$$P(C_0) = 1 - P(C_1)$$

$$P(C_1|X) = \frac{1}{1+e^{-w^T x}}$$

$$\therefore P(C_0|X) = 1 - P(C_1|X)$$

$$= 1 - \frac{1}{1+e^{-w^T x}}$$

$$= \frac{e^{-w^T x}}{1+e^{-w^T x}}$$

also, $\log\left(\frac{P(C_1|x)}{P(C_0|x)}\right) = \log\left(\frac{\frac{1}{1+e^{-w^Tx}}}{\frac{e^{-w^Tx}}{1+e^{-w^Tx}}}\right)$

$$= \log\left(\frac{1}{e^{-w^Tx}}\right)$$

$$= \log\left(e^{w^Tx}\right)$$

$$= w^Tx$$

Q3) The data with labels is given as:

$\{(x_i, c_i)\}$ where $x_i \in R^d$, $c_i \in [0, 1]$

~~also~~, thus, $y_i = w^Tx_i + w_0$

$$\sigma(y_i) = \frac{1}{1+e^{-y_i}} = \begin{cases} 1 & > 0.5 \\ 0 & < 0.5 \end{cases}$$

$\therefore$ according to the logistic regression, if $P(x_i)$ is probability of x_i,

$$P(x_i) = \frac{1}{1+e^{-(w^Tx+w_0)}}$$

1) Likelihood $L(w, w_0) = \prod_{i=1}^{N} p(x_i)^{c_i} \left(1 - p(x_i)\right)^{1-c_i}$

$$\therefore L(w, w_0) = \prod_{i=1}^{N} \left[\frac{1}{1+e^{-(w^T x_i + w_0)}}\right]^{c_i} \left[\frac{e^{-(w^T x_i + w_0)}}{1+e^{-(w^T x_i + w_0)}}\right]^{1-c_i}$$

2) <u>Log likelihood</u>

$$\log(L(w, w_0))$$

$$= \sum_{i=1}^{N} c_i \log\left[\frac{1}{1+e^{-(w^T x_i + w_0)}}\right] + (1-c_i)\log\left[\frac{e^{-(w^T x_i + w_0)}}{1+e^{-(w^T x_i + w_0)}}\right]$$

$$= \sum_{i=1}^{N} c_i \left[\log \frac{1}{1+e^{-(w^T x_i + w_0)}} - \log\left(\frac{e^{-(w^T x_i + w_0)}}{1+e^{-(w^T x_i + w_0)}}\right)\right] + \log\left[\frac{e^{-(w^T x_i + w_0)}}{1+e^{-(w^T x_i + w_0)}}\right]$$

$$= \sum_{i=1}^{N} c_i \log\left(e^{(w^T x_i + w_0)}\right) - \log\left[\frac{1 + e^{-(w^T x_i + w_0)}}{e^{-(w^T x_i + w_0)}}\right]$$

$$= \sum_{i=1}^{N} c_i \log\left(e^{(w^T x_i + w_0)}\right) - \log\left[1 + e^{(w^T x + w_0)}\right]$$

$$= \sum_{i=1}^{N} c_i (w^T x_i + w_0) - \log\left[1 + e^{(w^T x_i + w_0)}\right]$$

$\therefore$ Proved

3) From previous,

$$\log(L(w, w_0)) = \sum_{i=1}^{N} c_i (w^T x_i + w_0) - \log\left(1 + e^{(w^T x_i + w_0)}\right]$$

$$\frac{\partial L(w, w_0)}{\partial w_0} = \sum_{i=1}^{N} c_i - \frac{\partial}{\partial w_0} \log\left(1 + e^{(w^T x_i + w_0)}\right)$$

$$= \sum_{i=1}^{N} c_i - \frac{1}{1 + e^{(w^T x_i + w_0)}}$$

$$\frac{\partial}{\partial w_0}\left[1 + e^{(w^T x_i)} \cdot e^{(w_0)}\right]$$

$$= \sum_{i=1}^{N} c_i - \frac{e^{(w^T x_i + w_0)}}{1 + e^{(w^T x_i + w_0)}}$$

$$\frac{\partial L(w, w_0)}{\partial w} = \sum_{i=1}^{N} c_i x_i - \frac{1}{1 + e^{(w^T x_i + w_0)}}$$

$$\frac{\partial}{\partial w} \left[1 + e^{(w^T x_i + w_0)} \right]$$

$$= \sum_{i=1}^{N} c_i x_i - \frac{x_i e^{(w^T x_i + w_0)}}{1 + e^{(w^T x_i + w_0)}}$$