Shivam Jindal
2020125

Machine Learning
CSE343
Assignment-1

Section-A

Quest 1.

Ans- (a) we have been given a linear regression model and we need to prove that the least square fit line always passes through the point $(\bar{X}, \bar{Y})$ where $\bar{X}$ and $\bar{Y}$ represent the A.M. of independent variables and dependent variables respectively.

Equation for linear regression (model) is:

$$y_i = w_1 x_i + b$$ where b is bias and w_1 is weight

we need to prove: $\bar{Y} = w_1 \bar{X} + b$

Now, cost function is:

$$J(w) = \underset{w}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} (w_1 x_i + b - y_i)^2$$

In the case of least square fit, cost function will be minimum, so we will minimise $J(w)$ w.r.t. w_1, b.

i.e. $\frac{d(J(w))}{dw_1} = 0$ and $\frac{d(J(w))}{db} = 0$

$$\frac{d(J(\omega))}{d(\omega_1)} = \frac{2}{n} \sum_{i=1}^{n} (\omega_1 x_i - y_i + b)\, x_i = 0$$

and $\frac{d(J(\omega))}{d(b_0)} = 0$

$$\frac{d(J(\omega))}{db_0} = \frac{2}{n} \sum_{i=1}^{n} (\omega_1 x_i - y_i + b) = 0$$

i.e. $\sum_{i=1}^{n} (\omega_1 x_i - y_i + b) = 0$

$$\sum_{i=1}^{n} (\omega_1 x_i - y_i) + bn = 0$$

$$bn = -\sum_{i=1}^{n} (\omega_1 x_i - y_i)$$

$$\boxed{b = \frac{-1}{n} \sum_{i=1}^{n} (\omega_1 x_i - y_i)}$$

$$b = \frac{-1}{n}\left[\omega_1 (x_1 + x_2 + x_3 - - - x_n) - (y_1 + y_2 + y_3 - - y_n)\right]$$

$$b = -\frac{\omega_1 (x_1 + x_2 + x_3 - - x_n)}{n} + \frac{(y_1 + y_2 - - y_n)}{n}$$

i.e. $\boxed{\overline{Y} = \omega_1 \overline{X} + b}$

$\overline{Y}$ and $\overline{X}$ are the arithmetic means.
Since $(\overline{X}, \overline{Y})$ satisfies linear regression model.

Hence the point $(\overline{X}, \overline{Y})$ lies on it.

Quest 1.

Ans—(b) let $r_{x,y}$ denotes the correlation coefficient between x and y.

Since, it is given that $r_{x,z}$ and $r_{y,z}$ have high correlation, that means $r_{x,z}$ and $r_{y,z} > 0$.

Now, we need to estimate the correlation between x and y i.e. $r_{x,y}$.

By using partial correlation formula:

$$r_{x,y.z} = \frac{r_{x,y} - r_{x,z} \times r_{yz}}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}}$$

$$\boxed{r_{x,y} = r_{x,z} \cdot r_{y,z} + r_{x,y.z}\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}}$$

Since $-1 \leq r_{x,y.z} \leq 1$

So, we can write:

$$r_{x,y} = r_{x,z} \cdot r_{y,z} \pm \sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}$$

So, $r_{x,y} \in \left[r_{x,z} \cdot r_{y,z} - \sqrt{1-r_{zx}^2}\sqrt{1-r_{yz}^2},\; r_{x,z} \cdot r_{y,z} + \sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}\right]$

To take an example, let us take $r_{x,z} = 0.8$ and $r_{y,z} = 0.85$, for strong correlation.

Now, $r_{x,z} \cdot r_{y,z} = 0.68$

$\sqrt{1-r_{xz}^2} = 0.6$ and $\sqrt{1-r_{yz}^2} = 0.526$

$\sqrt{1-r_{xz}^2} \cdot \sqrt{1-r_{yz}^2} = 0.316$

$\Rightarrow r_{x,y} \in [0.68 - 0.316,\ 0.68 + 0.316]$

$r_{x,y} \in [0.36, 0.996]$

Since, we can see that $r_{x,y}$ can be as low as 0.36, for which we cannot say that strong correlation holds between x and y.

So, if two variables have a high correlation with a third variable, we cannot definitely say that they will also be highly correlated.

Ans- (c) We have to prove weak law of large numbers (LLN).

So, basically we can take a sequence $x_1, x_2, x_3, x_4 --- x_n$ of independent and identical random variables.

let expected value of each random variable is α.

i.e. $E(X_1) = E(X_2) --- E(X_n) = \alpha$.

α is the population mean.

we need to prove

$$\lim_{n \to \infty} \frac{x_1 + x_2 --- x_n}{n} \longrightarrow \alpha$$

let $X = x_1 + x_2 + x_3 --- x_n$

$$\overline{X} = \frac{x_1 + x_2 --- x_n}{n} = \text{Sample mean}$$

$\lim_{n \to \infty} \overline{X} \longrightarrow \alpha$, can be rewritten

as $\lim_{n \to \infty} P(|\overline{x} - \alpha|) \geq \epsilon) \longrightarrow 0$

for $\epsilon > 0$.

Now, by Chebyshev's inequality:

$$\boxed{P(|\overline{x} - \alpha| \geq \epsilon) \leq \frac{Var(\overline{X})}{\epsilon^2}}$$

$$\text{var}(\bar{x}) = \text{Var}\left(\frac{x_1 + x_2 - - - x_n}{n}\right)$$

Since all $x_1, x_2 - - x_n$ are independent and identical (iid).

Therefore, $\text{Var}(\bar{x}) = \frac{1}{n^2} \text{Var}(x_1 + x_2 - - x_n)$

$$= \frac{1}{n^2}\left(\text{Var}(x_1) + \text{Var}(x_2) - - \text{Var}(x_n)\right)$$

$$\text{Var}(x_i) = \sigma^2$$

$$\text{Var}(\bar{x}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\frac{\text{Var}(\bar{x})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

$$\lim_{n \to \infty} \therefore P(|\bar{x} - \alpha| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

$$\Downarrow$$

$$\lim_{n \to \infty} P(|\bar{x} - \alpha| \geq \epsilon) \leq 0 \quad (n \to \infty)$$

But probability ~~can't~~ can't be negative

So, $P(|\bar{x} - \alpha| \geq \epsilon) = 0$

Hence, we can say that $\bar{x} - \alpha < \epsilon$ and as $\epsilon \to 0$, so

$$\lim_{n \to \infty} (\bar{x} - \alpha) \to 0 \Rightarrow \lim_{n \to \infty} \bar{x} \to \alpha$$

Hence Proved.

Date

Pseudo code :

let us take range of numbers from 1 to 100.
So, our population mean $= \frac{5050}{100} = 50.5$

let us make an arr for sample data

```
arr = []
For i in range (0, 101):
    for j in range (i, 101, 4):
        if (!arr contains j):
            arr.add(j)

mean = mean(arr).
```

This mean is the sample mean. If we check sample mean at every iteration, we will see that gradually sample mean is getting more and more closer to population mean. If our dataset is very large, then sample mean will be equal to population mean.

Ans - (d) We have to derive Maximum A Posteriori (MAP) solⁿ for linear regression.

Model for linear regression:

$$y = w^T x + \epsilon$$

Since $\epsilon \sim Normal(0, \sigma^2)$

$y/x \sim Normal(w^T x, \sigma^2)$

$$\therefore P(y_i/x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w^T x_i - y_i)^2}{2\sigma^2}}$$

Now, $w = \underset{w}{argmax}\ P(w / y_1, x_1, y_2, x_2 - - y_n, x_n)$

By Bayes's Rule

$$w = \underset{w}{argmax} \frac{P(y_1, x_1, y_2, x_2 - - - y_n, x_n / w) P(w)}{P(y_1, x_1, y_2, x_2 - - - y_n, x_n)}$$

Now, we can ignore the denominator part.

$$w = \underset{w}{argmax}\ P(y_1, x_1, y_2, x_2 - - y_n, x_n / w) P(w)$$

Now, we will assume (weights are gaussian)

$w \sim Normal(0, \alpha^2)$

$$P(w) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{1}{2\alpha^2}(w-0)^2}$$

Page No.
Date

$$P(w) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{w^2}{2\alpha^2}}$$

Since, we can write,

$$P(y_1, x_1, \text{---} y_n, x_n / w) = \prod_{i=1}^{n} P(y_i, x_i / w)$$

and $P(y, x/w) = P(y/x, w) \, P(x/w)$

$$\prod_{i=1}^{n} P(y_i, x_i / w) = \prod_{i=1}^{n} P(y_i / x_i, w) \, P(x_i / w)$$

$$w = \underset{w}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i / x_i, w) P(x_i / w) \, P(w)$$

$$= \underset{w}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i / x_i, w) \, P(x_i) \, P(w)$$

Since $\underset{w}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i / x_i, w) = \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log P(y_i / x_i, w)$

and $P(x_i)$ is constant

$$w = \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log P(y_i / x_i, w) + \log P(w)$$

$$P(y_i / x_i, w) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(w^T x_i - y_i)^2}{2\sigma^2}}$$

$$P(w) = \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{w^2}{2\alpha^2}}$$

$$w = \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \log\left(e^{-\frac{(w^T x_i - y_i)^2}{2\sigma^2}}\right)$$

$$+ \log\left(\frac{1}{\sqrt{2\pi\alpha^2}}\right) + \log\left(e^{\frac{-w^2}{2\alpha^2}}\right)$$

(constant terms removed)

$$w = -\underset{w}{\operatorname{argmax}}\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{w^2}{2\alpha^2}\right)$$

$$\boxed{w = \underset{w}{\operatorname{argmin}}\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(w^T x_i - y_i)^2 + \frac{w^2}{2\alpha^2}\right)}$$