RANJAN KUMAR - 113262786
SHUBHAM AGRAWAL - 113166701
AMEYA SANKHE - 113219562
PRATIK NAGELIA - 114122204

CSE 544, Spring 2021: Probability and Statistics for Data Science

Assignment 6: Bayesian Inference and Regression Due: 05/06, 1:15pm, via Blackboard

(6 questions, 70 points total)

I/We understand and agree to the following:

(a) Academic dishonesty will result in an 'F' grade and referral to the Academic Judiciary.

(b) Late submission, beyond the 'due' date/time, will result in a score of 0 on this assignment.

(write down the name of all collaborating students on the line below)

RANJAN KUMAR, SHUBHAM AGRAWAL, AMEYA SANKHE, PRATIK NAGELIA

1. Posterior for Normal **(Total 10 points)**

Let $X_1, X_2, \ldots, X_n$ be distributed as Normal(θ, σ²), where σ is assumed to be known. You are also given that the prior for θ is Normal(a, b²).

(a) Show that the posterior of θ is Normal(x, y²), such that: (6 points)

$x = \frac{b^2\bar{X}+se^2a}{b^2+se^2}$ and $y^2 = \frac{b^2se^2}{b^2+se^2}$; where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $se^2 = \sigma^2/n$.

(Hint: less messier if you ignore the constants, but please justify why you can ignore them)

(b) Compute the (1-α) posterior interval for θ. (4 points)

(a) Ans. Posterior of $\theta \propto$ Likelihood (θ) × Prior (θ) — ①

Likelihood (θ) : $f(X/\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}\frac{1}{2}\left(\frac{x_i-\theta}{\sigma}\right)^2} \quad — ②$$

Prior $(\theta) = f(\theta) = \frac{1}{b\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\theta-a}{b}\right)^2}$ — ③

From Equation ①, we get :-

$$f(\theta/X) \propto \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \left(\frac{1}{b\sqrt{2\pi}}\right) e^{\left\{-\frac{1}{2}\left(\frac{\theta-a}{b}\right)^2 - \sum_{i=1}^{n}\frac{1}{2}\left(\frac{x_i-\theta}{\sigma}\right)^2\right\}}$$

{On removing the constants that is there since we are dealing with proportionality, only the terms where θ is there will be relevant}

So $f(\theta/X) \propto e^{-\frac{1}{2}\underbrace{\left[\left(\frac{\theta-a}{b}\right)^2 + \sum_{i=1}^{n}\left(\frac{X_i-\theta}{\sigma}\right)^2\right]}_{A}}$

Let $f(\theta/x) \propto e^{-\frac{1}{2}t}$

$$t = \left(\frac{\theta - a}{b}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i - \theta}{\sigma}\right)^2$$

$$= \frac{\theta^2 - 2a\theta + a^2}{b^2} + \frac{\sum x_i^2 - 2\sum_{i=1}^{n} x_i\theta + \sum_{i=1}^{n}\theta^2}{\sigma^2}$$

On Removing all the constant terms which does not involve θ, we get :-

$$t \propto \frac{\sigma^2\theta^2 - 2a\theta\sigma^2 + nb^2\theta^2 - 2n\bar{x}\theta b^2}{\sigma^2 b^2} \quad \left\{\text{using } \sum_{i=1}^{n} x_i = n\bar{x}\right\}$$

$$t \propto \frac{\theta^2(\sigma^2 + nb^2) - 2(\sigma^2 a + n\bar{x}b^2)\theta}{\sigma^2 b^2}$$

On dividing the numerator and denominator by $\sigma^2 + nb^2$,

$$t \propto \frac{\theta^2 - 2\theta\left(\frac{\sigma^2 a + nb^2\bar{x}}{\sigma^2 + nb^2}\right)}{\frac{\sigma^2 b^2}{\sigma^2 + nb^2}}$$

On adding and subtracting $\frac{\sigma^2 a + nb^2\bar{x}}{\sigma^2 + nb^2}$ in the above equation

we get :- $$t \propto \frac{\left[\theta - \left(\frac{\sigma^2 a + nb^2\bar{x}}{\sigma^2 + nb^2}\right)\right]^2}{\left(\frac{\sigma^2 b^2}{\sigma^2 + nb^2}\right)}$$

Therefore,

$$f(\theta/x) \propto e^{-\frac{1}{2}\left[\frac{\left(\theta - \frac{\sigma^2 a + n b^2 \bar{x}}{\sigma^2 + n b^2}\right)}{\frac{\sigma^2 b^2}{\sigma^2 + n b^2}}\right]}$$

On adjusting the constant of above proportionality, we get that $f(\theta/X)$ follows a Normal Distribution with mean $X = \frac{\sigma^2 a + n b^2 \bar{x}}{\sigma^2 + n b^2}$

and standard deviation, $y^2 = \frac{\sigma^2 b^2}{\sigma^2 + n b^2}$

Let $se^2 = \frac{\sigma^2}{n}$, so $X = \frac{\sigma^2 a + n b^2 \bar{x}}{\sigma^2 + n b^2}$

or, $X = \frac{\frac{\sigma^2 a}{n} + b^2 \bar{x}}{\frac{\sigma^2}{n} + b^2} = \frac{se^2 a + b^2 \bar{x}}{se^2 + b^2}$

{Proved}

Now $y^2 = \frac{\sigma^2 b^2}{\sigma^2 + n b^2} = \frac{\frac{\sigma^2}{n} \cdot b^2}{\frac{\sigma^2}{n} + b^2} = \frac{se^2 b^2}{se^2 + b^2}$

{Proved}

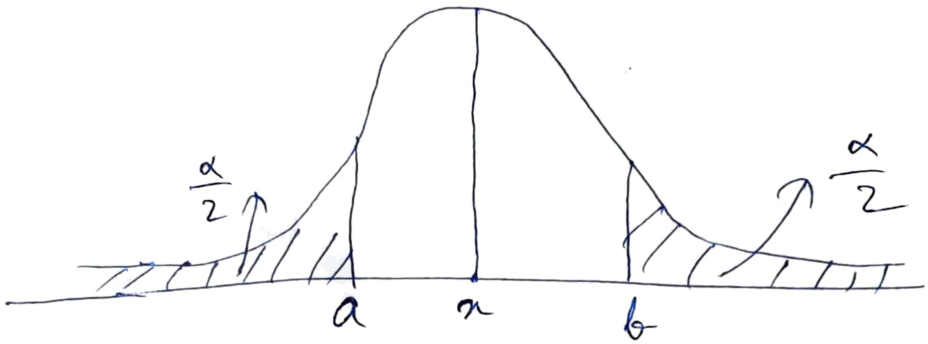
Therefore, Posterior for Normal = $f(\theta/X)$

= Normal (X, y^2)

(b) Ans. Let $\mathcal{D} = \{x_1, x_2, x_3 - - - - x_n\}$

We want interval $[a, b]$ such that

$$Pr(\theta \in [a, b]) \geq 1 - \alpha$$



$$Pr(\theta < a / \mathcal{D}) = \frac{\alpha}{2} \quad \text{--- (1)}$$

$$Pr(\theta > b / \mathcal{D}) = \frac{\alpha}{2} \quad \text{--- (2)}$$

On using the result from part a, and converting it to standard normal by subtracting the mean x and dividing by the standard deviation y, we get :-

$$Pr\left(\theta < \frac{a}{\mathcal{D}}\right) = Pr\left(\frac{\theta - x}{y} < \frac{a - x}{y} \Big| \mathcal{D}\right) = \frac{\alpha}{2}$$

$$\Rightarrow Pr\left(Z < \frac{a - x}{y}\right) = \frac{\alpha}{2} \quad \text{- (3)}$$

Z is the standard normal distribution.

Since $Pr\left(Z < -Z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ --- (4)

From (3) & (4), we get :- $\frac{a - x}{y} = -Z_{\frac{\alpha}{2}}$

or, $\boxed{a = x - y Z_{\alpha/2}}$

$$Pr\left(\theta > \frac{b}{\beta}\right) = \frac{\alpha}{2}$$

Converting to the standard normal by subtracting the mean x and dividing by standard deviation y we get:-

$$Pr\left(\frac{\theta - x}{y} > \frac{b - x}{y} \,\middle|\, \beta\right) = \frac{\alpha}{2}$$

$$Pr\left(Z > \frac{b-x}{y}\right) = \frac{\alpha}{2} \quad \text{— (5)}$$

we know that:-

$$Pr\left(Z \leq Z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$

or, $\frac{\alpha}{2} = 1 - Pr\left(Z \leq Z_{\frac{\alpha}{2}}\right)$

or, $\frac{\alpha}{2} = Pr\left(Z > Z_{\frac{\alpha}{2}}\right)$ — (6)

from (5) & (6), we get:-

$$\frac{b - x}{y} = Z_{\frac{\alpha}{2}}$$

$$\Rightarrow \boxed{b = x + y\, Z_{\frac{\alpha}{2}}}$$

$(1-\alpha)$ Posterior interval for $\theta = [a, b] =$

$$= \left[x - y z_{\frac{\alpha}{2}} , x + y z_{\frac{\alpha}{2}} \right]$$

On Putting the value of x and y as obtained in part 1a, we get :-

$(1-\alpha)$ Posterior Interval for θ is

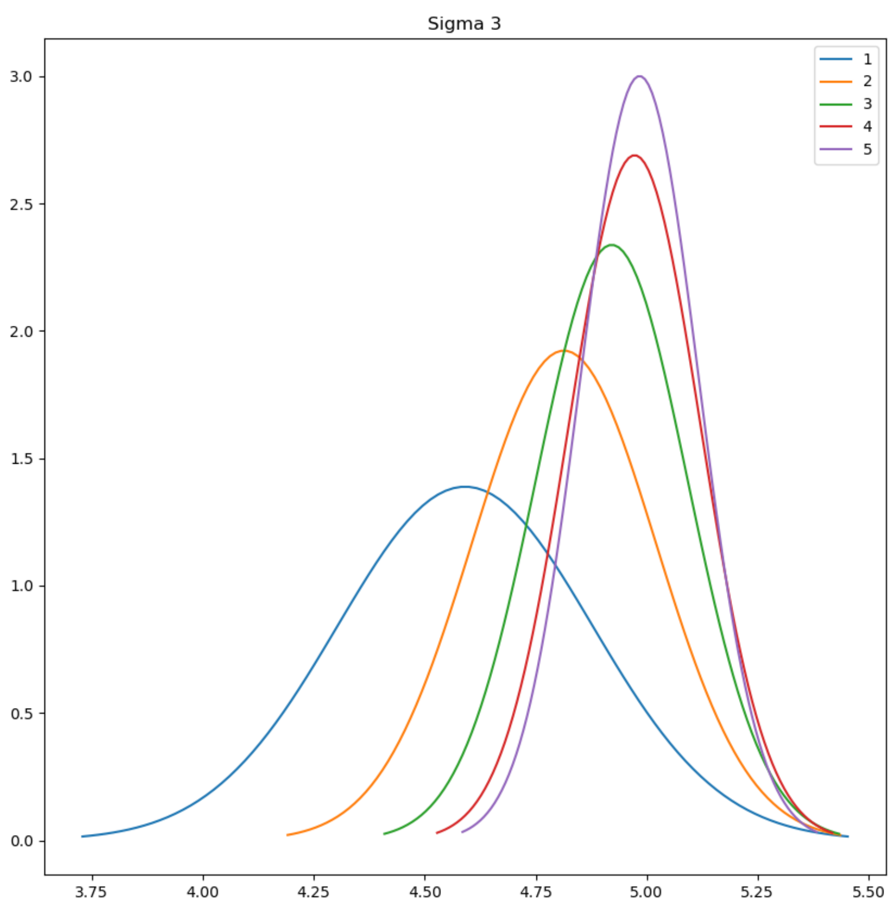
$$\left[\frac{b^2 \bar{X} + se^2 a}{b^2 + se^2} - z_{\frac{\alpha}{2}} \left(\frac{b\, se}{\sqrt{b^2 + se^2}} \right), \right.$$

$$\left. \frac{b^2 \bar{X} + se^2 a}{b^2 + se^2} + z_{\frac{\alpha}{2}} \left(\frac{b \cdot se}{\sqrt{b^2 + se^2}} \right) \right]$$

2)

a) sigma 3

	x	y_squared
0	4.590762	0.082569
1	4.813524	0.043062
2	4.921257	0.029126
3	4.972837	0.022005
4	4.983966	0.017682

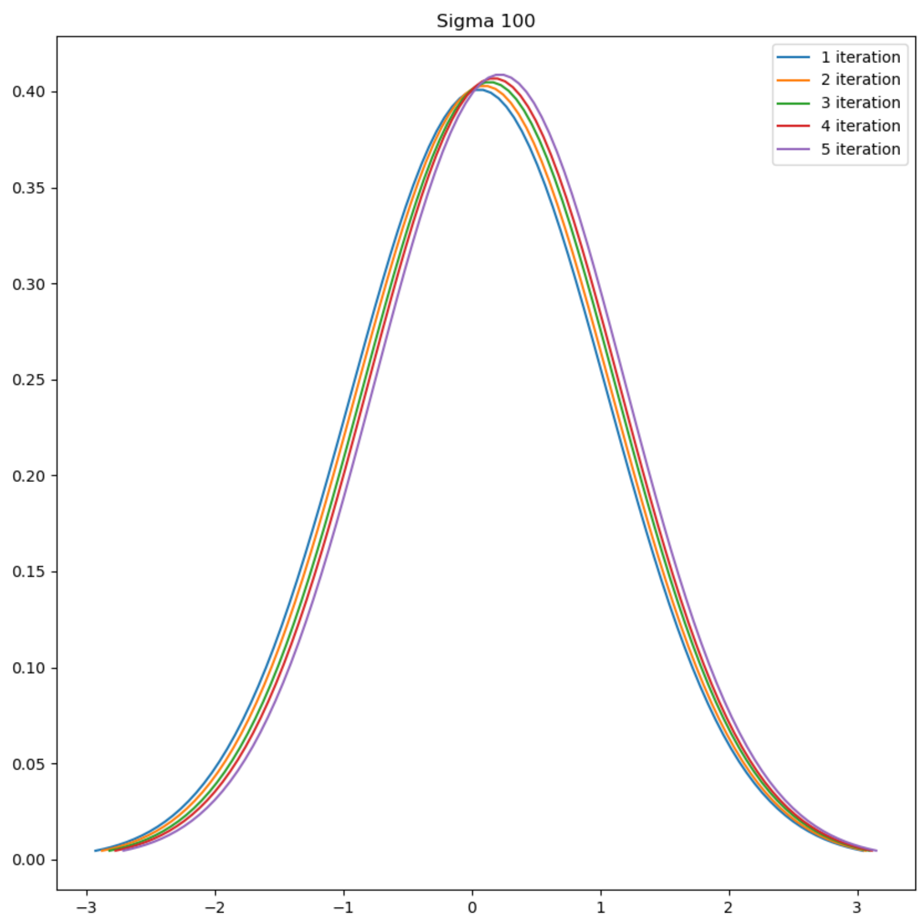


As we get more data points, the variance decreases and it becomes more confident and converges to value 5

****

2 b)
sigma 100

	x	y_squared
0	0.058716	0.990099
1	0.095009	0.980392
2	0.138226	0.970874
3	0.171219	0.961538
4	0.218918	0.952381



With higher variance, there is not much difference between the distributions
2c)
When the variance is less, the mean of the distribution is closer to the MLE
When the variance is more, the mean of the distribution is closer to the mean of the prior

3. Regression Analysis **(Total 7 points)**

Assume Simple Linear Regression on *n* sample points (Y_1, X_1), (Y_2, X_2), ..., (Y_n, X_n); that is, $Y = \beta_0 + \beta_1 X + \varepsilon_i$, where $E[\varepsilon_i] = 0$.

(a) Using the estimates of β derived in class, show that:

$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$ and $\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X}$, where $\bar{X} = (\sum_{i=1}^{n} X_i)/n$ and $\bar{Y} = (\sum_{i=1}^{n} Y_i)/n$. (2 points)

(b) Show that the above estimators, given X_is, are unbiased (Hint: Treat X's as constants) (5 points)

Q3.(a) Given, $Y = \beta_0 + \beta_1 X + \varepsilon_i$ ...... ①

where ε_i is the errors.

Also, $E[\varepsilon_i] = 0$

We know from lecture, $Y_i \mid X_i = \beta_0 + \beta_i X_i + \varepsilon_i$

① can be written as

Now taking Expectation on both sides.

$$E[Y_i \mid X_i] = E[\beta_0 + \beta_1 X_i + \varepsilon_i] \stackrel{LOE}{=} E[\beta_0] + E[\beta_1 X_i] + \cancel{E[\varepsilon_i]}^{0}$$

[Since β_0 & $\beta_1 X_i$ are constants]

$$E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$$

$$\therefore \hat{Y_i} = E[Y_i \mid X_i] = \hat{\beta_0} + \hat{\beta_1} X_i$$

The ~~sa~~ residual can be written as

$$\hat{\varepsilon_i} = Y_i - \hat{Y_i} = Y_i - (\hat{\beta_0} + \hat{\beta_1} X_i)$$

Sum of Squared Error :- SSS

$$S = \sum_{i=1}^{n} (\hat{\varepsilon_i})^2$$

To minimize $\hat{\beta_0}$ & $\hat{\beta_1}$, we need to take partial derivatives.

$$S = \sum_{i=1}^{n} (Y_i - \hat{Y_i})^2 = \sum_{i=1}^{n} (Y_i - (\hat{\beta_0} + \hat{\beta_1} X_i))^2$$

Taking Partial Derivatives of S w.r.t. $\hat{\beta}_0$.

$$\frac{\partial S}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} 2(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))(0-1) = 0$$

$$0 = -\sum_{i=1}^{n} 2(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i) = n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_i$$

Dividing by n on both sides,

$$\frac{\Sigma Y_i}{n} = \frac{n\hat{\beta}_0}{n} + \frac{\hat{\beta}_1 \Sigma X_i}{n}$$

where $\frac{\Sigma Y_i}{n} = \bar{Y}$

$$\boxed{\bar{Y} = \hat{\beta}_0 + \bar{X}\hat{\beta}_1} \qquad \boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}} \quad \text{--- (2)}$$

$\frac{\Sigma X_i}{n} = \bar{X}$

Taking partial Derivative of S w.r.t. $\hat{\beta}_1$ and put is equal to 0

$$\frac{\partial S}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))(-X_i) = 0$$

$$\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0 \quad \text{.... (3)}$$

Put (2) ..... in (3), i.e. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ we get.

$$\sum_{i=1}^{n} Y_i x_i - (\bar{Y} - \hat{\beta}_1 \bar{x})\sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$\sum_{i=1}^{n} Y_i X_i - \bar{Y}\sum_{i=1}^{n} X_i + \hat{\beta}_1 \bar{X} \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} Y_i X_i - \bar{Y}\sum_{i=1}^{n} X_i = \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i\right)$$

$$\sum_{i=1}^{n} X_i Y_i - \bar{Y}\sum_{i=1}^{n} x_i - \bar{X}\sum_{i=1}^{n} Y_i + \bar{X}\sum Y_i = \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X} \sum X_i^n\right)$$

[Adding and Subtracting $\bar{X}\Sigma Y_i$ on LHS)

And putting $\sum_{i=1}^{n} Y_i = n\bar{Y}_i$

$$\sum_{i=1}^{n} X_i Y_i - \bar{X}\sum_{i=1}^{n} Y_i - \bar{Y}\sum_{i=1}^{n} X_i + n\bar{X}\bar{Y}$$

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i\right)$$

Writing $n\bar{X}\bar{Y} = \sum_{i=1}^{n} \bar{X}\bar{Y}$

we get LHS $= \sum_{i=1}^{n} Y_i X_i - \bar{X}\sum_{i=1}^{n} Y_i - \bar{Y}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \bar{X}\bar{Y}$

$$= \sum_{i=1}^{n} \left[X_i Y_i - \bar{X} Y_i - \bar{Y} X_i + \bar{X}\bar{Y}\right]$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Now lets see RHS.

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i\right)$$

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - \bar{X}\sum_{i=1}^{n} X_i - \bar{X}\sum_{i=1}^{n} X_i + \bar{X}\sum_{i=1}^{n} X_i\right)$$

[Adding & Subtracting $\bar{X}\sum_{i=1}^{n} X_i$]

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + \bar{X}\sum_{i=1}^{n} X_i\right)$$

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}\bar{X}\right) \quad \left[\because \sum_{i=1}^{n} X_i = n\bar{X}\right]$$

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \bar{X}^2\right)$$

$$= \hat{\beta}_1 \left[\sum_{i=1}^{n} \left(X_i^2 - 2\bar{X}X_i + \bar{X}^2\right)\right] \quad \left[\text{Taking out } \sum_{i=1}^{n}\right]$$

$$= \hat{\beta}_1 \left(\sum_{i=1}^{n} (X_i - \bar{X})^2\right)$$

$\therefore$ Putting LHS = RHS, we get

$$\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \hat{\beta}_1 \left(\sum_{i=1}^{n} (X_i - \bar{X})^2\right)$$

$\therefore$ We get $\boxed{\hat{\beta_1} = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}}$

By $\boxed{\hat{\beta_0} = \bar{Y} - \hat{\beta_1} \bar{X}}$

(b) $\text{Bias}(\theta) = E[\theta] - \theta$

$$E[\hat{\beta_0}] = E[\bar{Y} - \hat{\beta_1} \bar{X}]$$

$$= E\left[\frac{\sum_{i=1}^{n} Y_i}{n} - \hat{\beta_1} \frac{\sum_{i=1}^{n} X_i}{n}\right] = E\left[\frac{\sum_{i=1}^{n} Y_i - \hat{\beta_1} \sum_{i=1}^{n} X_i}{n}\right]$$

(Taking n out of Expectation) $= \dfrac{1}{n} E\left[\sum_{i=1}^{n} (Y_i - \hat{\beta_1} X_i)\right]$

$$\stackrel{\text{LOE}}{=} \frac{\sum_{i=1}^{n} E[Y_i - \hat{\beta_1} X_i]}{n}$$ (Using Linearity of Expectation)

$$= \sum_{i=1}^{n} \left(\frac{E[Y_i] - E[\hat{\beta_1}] \cdot X_i}{n}\right)$$ Taking X_i as constant.

$\because$ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$E[Y_i] = \beta_0 + \beta_1 X_i$ .... (From Part A, & $E[\varepsilon_i] = 0$) ... (0)

$\therefore$

$$E[\hat{\beta_0}] = \frac{\sum_{i=1}^{n} (\beta_0 + \beta_1 X_i - E[\hat{\beta_1}] \cdot X_i)}{n} \quad \text{.... (1)}$$

To calculate this, we need the $E[\hat{\beta_1}]$.

From class, we know,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i Y_i) - n\bar{X}\bar{Y}}{\sum_{i=1}^{n}(x_i)^2 - n(\bar{x})^2}$$

Taking Expectation on both sides,

$$E[\hat{\beta}_1] = E\left[\frac{\Sigma(x_i Y_i) - n\bar{x}\bar{Y}}{\Sigma x_i^2 - n(\bar{x})^2}\right]$$

$$= \frac{E[\Sigma(x_i Y_i) - n\bar{x}\bar{Y}]}{\Sigma x_i^2 - n\bar{x}^2}$$

(Since Denominator is const)

$$\overset{LOV}{=} \frac{\Sigma E[x_i Y_i - n\bar{x}\bar{Y}]}{\Sigma x_i^2 - n\bar{x}^2} = \frac{\Sigma x_i E[Y_i] - n\bar{X}\bar{Y}}{\Sigma x_i^2 - n\bar{x}^2}$$

(Since, x_i is const. so is $n\bar{x}\bar{Y}$)

~~$= \Sigma\, x_i\, E[$~~ Putting $E[Y_i] = \beta_0 + \beta_1 X_i$ from Equation (0) we get.

$$= \frac{\Sigma\left(x_i(\beta_0 + \beta_1 x_i) - n\bar{x}\bar{Y}\right)}{\Sigma x_i - n\bar{x}^2}$$

$\left[\bar{Y} = \beta_0 + \beta_1\bar{X}\right]$ From class,

$$= \frac{\beta_0 \Sigma x_i + \beta_1 \Sigma x_i^2 - n\bar{x}(\beta_0 + \beta_1\bar{x})}{\Sigma x_i - n\bar{x}^2}$$

Putting $\Sigma x_i = n\bar{X}$, we get.

$$= \frac{\beta_0 \Sigma x_i + \beta_1 \Sigma x_i^2 - n\bar{x}\beta_0 - n\beta_1\bar{x}^2}{\Sigma x_i - n\bar{x}^2}$$

$$= \frac{\beta_0 \cancel{n\bar{x}} + \beta_1 \Sigma x_i^2 - \cancel{n\bar{x}\beta_0} - n\beta_1\bar{x}^2}{\Sigma x_i - n\bar{x}^2}$$

$$= \frac{\beta_1(\Sigma x_i^2 - n\bar{x}^2)}{(\Sigma x_i^2 - n\bar{x}^2)} = \beta_1$$

$\therefore E[\hat{\beta}_1] = \beta_1$. Hence $\hat{\beta}_1$ is unbiased.

Now putting the value of $E[\hat{\beta}_1]$ in equation (1) we get,

$$E[\hat{\beta_0}] = \frac{\sum_{i=1}^{n} \left(\beta_0 + \beta_1 X_i - E[\hat{\beta_1}] \cdot X_i\right)}{n}$$

$$= \frac{\sum_{i=1}^{n} \left(\beta_0 + \cancel{\beta_1 X_i} - \cancel{\beta_1 X_i}\right)}{n}$$

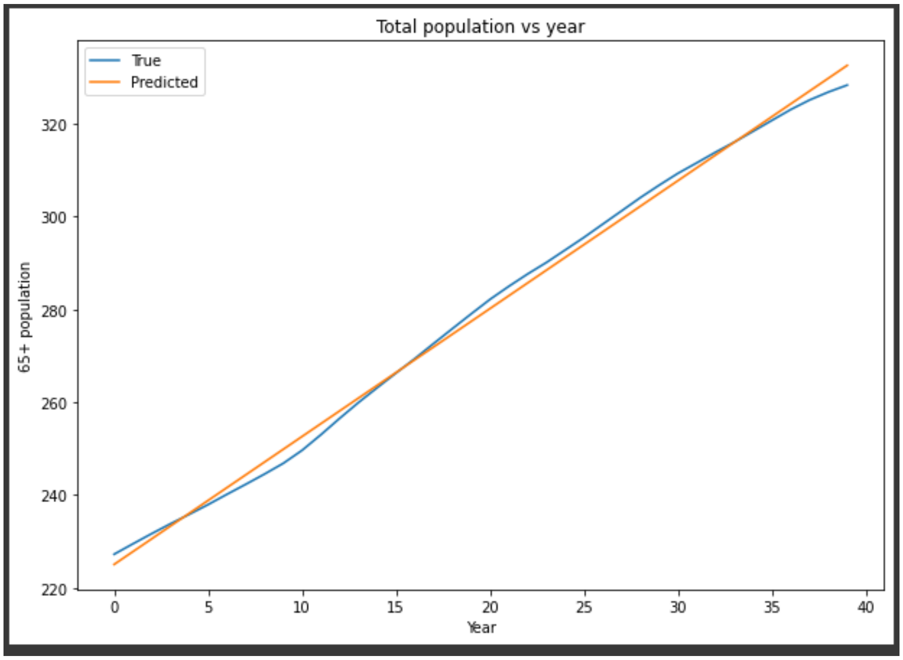
$$= \frac{\cancel{n}\beta_0}{\cancel{n}} = \beta_0 .$$

$\therefore E[\hat{\beta_0}] = \beta_0$.

Hence $\hat{\beta_0}$ is unbiased.

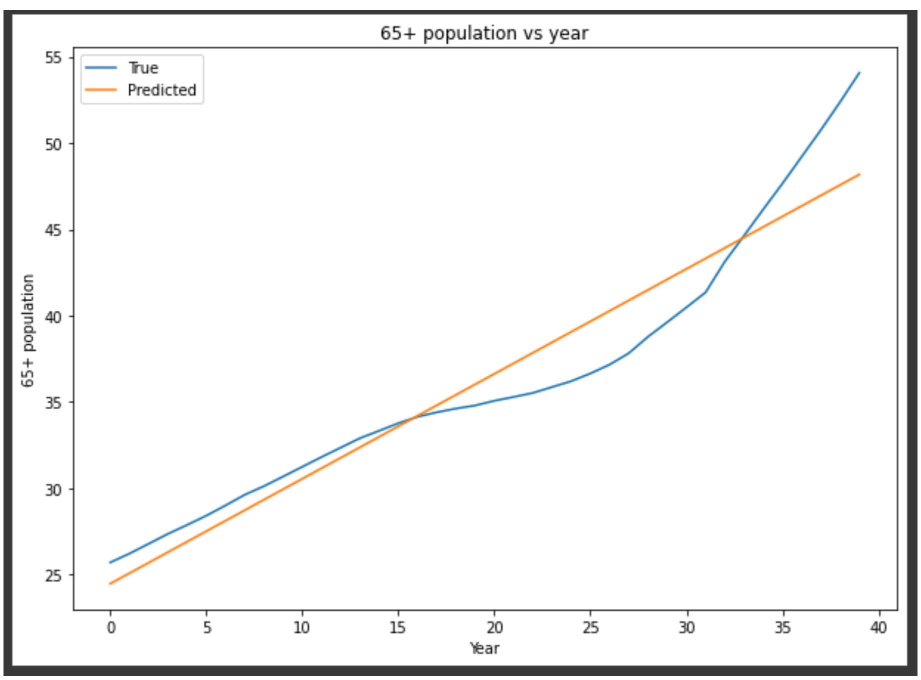
$\therefore$ Both $\hat{\beta_1}$ and $\hat{\beta_0}$ are unbiased. Hence Proved.

Q4 a)



SSE of Total Population: 123.8493735899118

Equation: y = -5234.856475245826 + 2.7575177477473503 * x



SSE of 65+ Population: 176.03538362528215

Equation: y = -1178.5613646417462 + 0.6075945051596631 * x

Q4 b)

Data From 1980 to 2018

SSE of 65+ Population: 137.69260032800028

Equation: y = -1131.0736031458225 + 0.5837632431175469 * x

prediction for 2060 in millions : 71.4786776763242

****

Data From 2008 to 2018

SSE of 65+ Population: 2.008761070738738

Equation: y = -2778.38941090817 + 1.4025382727268152 * x

prediction for 2060 in millions : 110.8394309090695

Q4 c)

Using Ratio Directly
Ratio of 65+ population: 0.16236192897081914

****

Population of 65+: 53.33536172727008

Total Population: 329.6888533636047

Ratio of 65+ population: 0.16177484068121645

****

True Ratio of 65+ population: 0.164646

Method 1 is more accurate and is closer to the True Ratio of 65+ population.

5
a)
Using Full Dataset

Beta
[[-0.00291067]
 [0.00323154]
 [0.01990962]
 [0.00057609]
 [0.02319267]
 [0.1308982]
 [0.05682043]]

Test SSE: 0.3164114091070942

b)
Using TOEFL, SOP, LOR

Beta
[[0.00388655]
 [0.04187385]
 [0.04825699]]

Test SSE: 0.6403887599866923

c)
Using GRE, GPA

Beta
[[-0.00410631]
 [0.23571159]]

Test SSE: 0.46380506786290315

d)
According to the above results, we can see that the part A which takes all the features has the lowest SSE value. Therefore, more features help us to predict better results.
Part C has better SSE than Part B, which shows that GRE, GPA are better features for prediction than TOEFL, SOP and LOR.

Overall, Part A gives the best prediction as it contains all the features.

6. Bayesian hypothesis testing **(Total 18 points)**

You are tired of studying probs and stats and have finally decided to give up your current life and turn to your one true passion – farming. Lucky for you, there is lot of farmland on Long Island, and you have your heart set on a particular farm that is available for purchase. However, you do not know whether the soil in the farm is good or not. Say the soil in the farm is a discrete random variable H and it can only take values in the set $\{0, 1\}$, where 0 represent good soil and 1 represents bad soil. We transform this as a hypothesis test as follows: $\boldsymbol{H_0}: H = 0$ and $\boldsymbol{H_1}: H = 1$. Let the prior probability $P(\boldsymbol{H_0}) = P(H = 0) = p$ and $P(\boldsymbol{H_1}) = P(H = 1) = 1 - p$. The water content in the soil depends upon the type of soil. If we assume water content to be a RV W, then $f_W(w|H = 0) = N(w; -\mu, \sigma^2)$ and $f_W(w|H = 1) = N(w; \mu, \sigma^2)$. To test which of the two hypotheses is correct, you take n samples of the soil from different patches of the farm and measure the water content metric of each sample; the resulting data sample set is $\boldsymbol{w} = \{w_1, w_2, w_3 \dots, w_n\}$. Assume that the samples are conditionally independent given the hypothesis/soil type.

(a) If we denote the hypothesis chosen as a RV C where $C \in \{0, 1\}$, then according to MAP (Maximum a posteriori), we have $C = \begin{cases} 0 & if\ P(H = 0|\boldsymbol{w}) \geq P(H = 1|\boldsymbol{w}) \\ 1 & otherwise \end{cases}$. This implies that the hypothesis H=0 is chosen (referring to C=0) when P(H=0|w) ≥ P(H=1|w). Derive a condition for choosing the hypothesis that soil in the farm is of type is 0, in terms of $p, \mu\ and\ \sigma$. **(4 points)**

(b) Write a python function **MAP_descision()** in a script named Q6_b.py, where your function takes as input (i) the list of observations $\boldsymbol{w}$, and (ii) the prior probability of $\boldsymbol{H_0}$, and returns the chosen hypothesis (value of C) according to the MAP criterion. Report the result for the 10 different instances of observations from the q6.csv dataset and for each prior probability p = [0.1, 0.3, 0.5, 0.8] for the value of (μ, σ^2) = (0.5, 1.0). Each column is one set of observations. **(10 points)**

Example output format:

For $P(H_0) = 0.1$, the hypotheses selected are :: 0 1 0 1 0 0 1 0 0 1
For $P(H_0) = 0.3$, the hypotheses selected are :: 1 1 0 1 1 0 0 0 0 1
For $P(H_0) = 0.5$, the hypotheses selected are :: 1 1 0 1 1 0 0 0 0 1
For $P(H_0) = 0.8$, the hypotheses selected are :: 1 1 0 1 1 0 0 0 0 1

(c) Denoting the hypothesis selected as a RV C where $C \in \{0, 1\}$, the average error probability via the MAP criterion is given by $\boldsymbol{AEP} = P(C = 0|H = 1)P(H = 1) + P(C = 1|H = 0)P(H = 0)$. Given the observations $\boldsymbol{w} = \{w_1, w_2, w_3 \dots, w_n\}$, derive $\boldsymbol{AEP}$ in terms of $\mu, \sigma, \Phi(\ \ )\ and\ p$. **(4 points)**

6 (a)

Given: $f_w(\omega | H=0) = N(\omega; -\mu, \sigma^2)$ where W is a Random Variable

$f_w(\omega | H=1) = N(\omega; \mu, \sigma^2)$

$P(H_0) = p$, $P(H_1) = 1-p$

Random Variable $C = \begin{cases} 0 & \text{if } P(H=0|w) \geq P(H=1|\omega) \\ 1 \end{cases}$

Goal: To derive a condition for choosing the hypothesis that soil in the farm is of type is 0

i.e. we choose $C = 0$, $H=0$ (good soil) iff

$P(H=0|w) \geq P(H=1|w)$ — eq. 1

From Bayes Theorem:

$$P(H=0|W) = \frac{P(W|H=0)\cdot P(H=0)}{P(W)} \quad \text{— eq}^n\text{. (2)}$$

$$P(H=1|W) = \frac{P(W|H=1)\cdot P(H=1)}{P(W)} \quad \text{— eq}^n\text{. (3)}$$

Putting the values of eq^n. (2) and (3) in (1), we get:

$$P(H=0|W) \geq P(H=1|W)$$

$$\frac{P(W|H=0)\cdot P(H=0)}{P(W)} \geq \frac{P(W|H=1)\cdot P(H=1)}{P(W)}$$

Since $P(W)$ ~~has~~ to be a positive value

$$\Rightarrow P(W|H=0)\cdot P(H=0) \geq P(W|H=1)\cdot P(H=1)$$

$$\Rightarrow p\cdot P(W|H=0) \geq (1-p)\,P(W|H=1) \quad \left[\begin{matrix} P(H_0)=p \\ P(H_1)=1-p \end{matrix}\right\}\text{given}\Big]$$

For sample set $W = \{w_1, w_2 \ldots w_n\}$

$$\Rightarrow p\cdot P(w_1, w_2 \ldots w_n|H=0) \geq (1-p)\cdot P(w_1, w_2 \ldots w_n|H=1)$$

As samples are conditionally independent

$$\Rightarrow p\cdot \prod_{i=1}^{n} P(w_i|H=0) \geq (1-p)\cdot \prod_{i=1}^{n} P(w_i|H=1)$$

Substituting given values for $P(w_i|H=0)$ and $P(w_i|H=1)$

$$\Rightarrow p\cdot \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{w_i+\mu}{\sigma}\right)^2} \geq (1-p)\prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{w_i-\mu}{\sigma}\right)^2}$$

$$\Rightarrow p\cdot e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(w_i+\mu)^2}{\sigma^2}} \geq (1-p)\cdot e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(w_i-\mu)^2}{\sigma^2}}$$

$$\Rightarrow \frac{e^{-\frac{1}{2}\cdot\frac{\sum_{i=1}^{n}(w_i+\mu)^2}{\sigma^2}}}{e^{-\frac{1}{2}\cdot\frac{\sum_{i=1}^{n}(w_i-\mu)^2}{\sigma^2}}} \geq \frac{(1-p)}{p}$$

$$\Rightarrow e^{\left(-\frac{\sum_{i=1}^{n}(w_i+\mu)^2}{2\sigma^2}+\frac{\sum_{i=1}^{n}(w_i-\mu)^2}{2\sigma^2}\right)} \geq \frac{(1-p)}{p}$$

$$\Rightarrow e^{\frac{\left(-\cancel{\sum_{i=1}^{n} w_i^2}-\cancel{\sum_{i=1}^{n}\mu^2}-2\sum_{i=1}^{n} w_i\mu+\cancel{\sum_{i=1}^{n} w_i^2}+\cancel{\sum_{i=1}^{n}\mu^2}-2\sum_{i=1}^{n} w_i\mu\right)}{2\sigma^2}} \geq \frac{(1-p)}{p}$$

$$\Rightarrow e^{\frac{-\cancel{4}^{2}\sum_{i=1}^{n} w_i\mu}{\cancel{2}\sigma^2}} \geq \frac{1-p}{p}$$

$$\Rightarrow e^{-2\sum_{i=1}^{n}\frac{w_i\mu}{\sigma^2}} \geq \frac{1-p}{p}$$

$$\Rightarrow e^{2\sum_{i=1}^{n}\frac{w_i\mu}{\sigma^2}} \leq \frac{p}{1-p}$$

Taking log on both sides

$$\frac{2\mu\sum_{i=1}^{n} w_i}{\sigma^2} \leq \ln\left(\frac{p}{1-p}\right)$$

$$\boxed{\therefore \sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right)}$$

Answer.

6)

b)

For P(H0) = 0.1, the hypotheses selected are :: 0 1 0 0 1 0 1 1 0 1
For P(H0) = 0.3, the hypotheses selected are :: 0 1 0 0 1 0 1 1 0 1
For P(H0) = 0.5, the hypotheses selected are :: 0 1 0 0 1 0 1 1 0 1
For P(H0) = 0.8, the hypotheses selected are :: 0 1 0 0 1 0 1 1 0 1

6(c) Given:

Average Error Probability $= P(C=0 \mid H=1)\, P(H=1) + P(C=1 \mid H=0) \cdot P(H=0)$

From 6(a), we choose $H=0$, iff

$$\sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)$$

we choose $H=1$, iff

$$\sum_{i=1}^{n} w_i > \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)$$

$$P(C=0 \mid H=1) = P\left(\left(\sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)\right) \Big| H=1\right)$$

Given: $f_W(w \mid H=0) \sim N(-n\mu, \sigma^2)$, $P(H=1) = 1-p$

$f_W(w \mid H=1) \sim N(n\mu, \sigma^2)$, $P(H=0) = p$

$$P(C=0 \mid H=1) = \Phi\left(\frac{\frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) - n\mu}{\sqrt{n\sigma^2}}\right) \quad \text{--- eq}^n \text{ (1)}$$

Since, $X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$

Similarly; $P(C=1 \mid H=0) = P\left(\left(\sum_{i=1}^{n} w_i > \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)\right) \Big| H=0\right)$

$$= 1 - \Phi\left(\frac{\frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) + n\mu}{\sqrt{n\sigma^2}}\right) \quad \text{--- eq}^n \text{ (2)}$$

$\therefore$ Average error probability

$$= (1-p) \cdot \Phi\left(\frac{\frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) - n\mu}{\sqrt{n\sigma^2}}\right) + p\left(1 - \Phi\left(\frac{\frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) + n\mu}{\sqrt{n\sigma^2}}\right)\right)$$

$$\Rightarrow e^{\left(\frac{-\sum_{i=1}^{n}(w_i+\mu)^2}{2\sigma^2} + \frac{\sum_{i=1}^{n}(w_i-\mu)^2}{2\sigma^2}\right)} \geq \frac{(1-p)}{p}$$

$$\Rightarrow e^{\frac{\left(-\cancel{\sum_{i=1}^{n} w_i^2} - \cancel{\sum_{i=1}^{n}\mu^2} - 2\sum_{i=1}^{n} w_i\mu + \cancel{\sum_{i=1}^{n} w_i^2} + \cancel{\sum_{i=1}^{n}\mu^2} - 2\sum_{i=1}^{n} w_i\mu\right)}{2\sigma^2}} \geq \frac{(1-p)}{p}$$

$$\Rightarrow e^{\frac{-\cancel{4}^{2}\sum_{i=1}^{n} w_i\mu}{\cancel{2}\sigma^2}} \geq \frac{1-p}{p}$$

$$\Rightarrow e^{-2\sum_{i=1}^{n}\frac{w_i\mu}{\sigma^2}} \geq \frac{1-p}{p}$$

$$\Rightarrow e^{2\sum_{i=1}^{n}\frac{w_i\mu}{\sigma^2}} \leq \cancel{\frac{1+p}{p}}\ \frac{p}{1-p}$$

Taking log on both sides

$$\frac{2\mu\sum_{i=1}^{n} w_i}{\sigma^2} \leq \ln\left(\frac{p}{1-p}\right)$$

$$\boxed{\therefore \sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right)}$$

Answer.