Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

1)

$$\sum_{i=1}^{m} (\hat{\omega} x_i + \hat{b} - y_i)^2$$

→ to find $\hat{\omega}$, $\hat{b}$ minimize the error

* find the derivative with respect to $\hat{\omega}$, $\hat{b}$ to find the values.

$$\left[E[x] = \frac{1}{m} \sum_{i=1}^{m} x_i \right]$$

$$\left[cov(x,y) = E[xy] - E[x]E[y] \right]$$

* $$\frac{d}{d\hat{\omega}} \sum_{i=1}^{m} (\hat{\omega} x_i + \hat{b} - y_i)^2 = 0$$

$$\sum_{i=1}^{m} 2(\hat{\omega} x_i + \hat{b} - y_i) * x_i = 0$$

$$\sum_{i=1}^{m} \hat{\omega} x_i^2 + \hat{b} x_i - x_i y_i = 0$$

$$= 2m\hat{\omega} E[x] + 2m\hat{b} E[x] - 2m E[xY] = 0$$

$$= \hat{\omega} E[x] + \hat{b} E[x] - E[xY] = 0$$

$$\hat{\omega} = \frac{E[xY] - \hat{b} E[x]}{E[x]} \dashrightarrow (1)$$

* $$\frac{d}{d\hat{b}} \sum_{i=1}^{m} (\hat{\omega} x_i + \hat{b} - y_i)^2$$

$$\sum_{i=1}^{m} 2(\hat{\omega} x_i + \hat{b} - y_i) = 0$$

$$m\hat{\omega} E[x] + m\hat{b} - m E[y] = 0$$

$$\hat{\omega} E[x] + \hat{b} - E[y] = 0$$

$$\hat{b} = E[y] - \hat{\omega} E[x] \dashrightarrow (2)$$

---> substitute (2) in (1)

$$\hat{\omega} = \frac{E[xy] - (E[y] - \hat{\omega} E[x]) E[x]}{E[x]}$$

$$\hat{\omega} = \frac{E[xy]}{E[x]} - E[y] + \hat{\omega} E[x]$$

$$\hat{\omega}(1 - E[x]) = \frac{E[xy]}{E[x]} - E[y]$$

$$\frac{E[xy] - E[x]E[y]}{E[x] - E[x]^2}$$

$$\boxed{\begin{aligned} \hat{\omega} &= \frac{E[xy] - E[x]E[y]}{E[x^2] - E[x]^2} \\ \hat{b} &= E[y] - \hat{\omega} E[x] \end{aligned}} \longrightarrow \frac{cov(x,y)}{Var(x)}$$

Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

(2)

$$y_i = w x_i + b + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2)$$

$$\left.\begin{aligned} E[\hat{w}] &= w \\ E[\hat{b}] &= b \end{aligned}\right] \rightarrow$$ argue that the estimators unbiased and what is the variance.

$$\hat{w} = \frac{E[xy] - E[x]E[y]}{E[x^2] - E[x]^2} = \frac{Cov(x,y)}{var(x)}$$

$$\hat{b} = E[y] - \hat{w} E[x]$$

$$\rightarrow E[\hat{w}] = \frac{\frac{\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})}{m}}{\frac{\sum (x_i - \bar{x})^2}{m}} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

$$E[\hat{w}] = \frac{\sum_{i=1}^{m} (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

$$E[\hat{w}] = \sum_{i=1}^{m} \frac{y_i (x_i - \bar{x}) - \bar{y}(x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

$$\sum_{i=1}^{m} \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})}{\sum_{i=1}^{m} (\bar{x}_i - \bar{x})^2} * \left(\sum_{i=1}^{m} w x_i + b \epsilon_i - \bar{y}\right)$$

$$E[\hat{w}] = \sum_{i=1}^{m} \frac{(x_i - \bar{x})}{(\bar{x}_i - \bar{x})^2} * \left(m w E[x] + bm \cancel{E[\epsilon]} - \cancel{E[\bar{y}]}\right)$$

0 0

$$= w \left(\frac{\sum_{i=1}^{m} (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} * \sum_{i=1}^{m} (x_i) \right)$$

$$\hookrightarrow \frac{\overline{x^2} - \bar{x}^2}{\overline{x^2} - \bar{x}^2} = 1$$

$$\boxed{E[\hat{w}] = w}$$

$$\rightarrow E[\hat{b}] = E[E[y]] - E[\hat{w}] * E[x]$$

$$= E[w\bar{x} + b + \epsilon] - E[\hat{w}] * E[x]$$

$$= \cancel{E[w]}\bar{x} + b + \cancel{E[\epsilon]} - \cancel{E[\hat{w}]} * \bar{x}$$

0

$$\boxed{E[\hat{b}] = b}$$

Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

(*) Finding the variance

$$\hat{\omega} = \sum_{i=1}^{m} \frac{(x_i - \bar{x}}{\sum (x_i - \bar{x})^2} (y - \bar{y})$$

$$\hat{\omega} = \frac{\sum_{i=1}^{m} (x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2} * ((\omega x_i + b + \epsilon_i) - \bar{y})$$

$$\hat{\omega} = \omega * \underbrace{\sum_{i=1}^{m} \frac{(x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2} * x_i}_{1} + \sum_{i=1}^{m} \frac{(x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2} * b + \sum_{i=1}^{m} \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} * \epsilon_i - \bar{y}$$

$$= \underbrace{\omega * 1 + \sum_{i=1}^{m} \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} b}_{\text{constant}} + \sum_{i=1}^{m} \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \epsilon_i - \bar{y}$$

$$Var(\hat{\omega}) = Var\left(\omega + \frac{\sum_{i=1}^{m} (x_i - \bar{x}) \epsilon_i}{\sum (x_i - \bar{x})^2}\right)$$

$$Var(\hat{\omega}) = \left(\frac{\sum_{i=1}^{m} (x_i - \bar{x})}{\sum_{i=1}^{m} (x_i - \bar{x})^2}\right)^2 * Var(\epsilon_i)$$

$$Var(\hat{\omega}) = \frac{Var(x) * \frac{\sigma^2}{m}}{(Var(x))^2} = \frac{\sigma^2}{m Var(x)}$$

$$\boxed{Var(\hat{\omega}) = \frac{\sigma^2}{m * Var(x)}}$$

$\rightarrow$ $\hat{b} = E[y] - \hat{\omega} E[x]$

$$Var(\hat{b}) = Var(E[y] - \hat{\omega} E[x])$$

$$= Var(E[y]) + E[x^2] Var(\hat{\omega}) - 2 Cov(\underbrace{E[y], \hat{\omega} E[x]}_{0})$$

$$= \frac{E[x^2] * \sigma^2}{m * Var(x)}$$

(3) with $var(x)$ as variance and $E[x]$ as expectation from the calculation and proof above we found that the var will actually be around the following

$$Var(\hat{\omega}) = \frac{\sigma^2}{m * Var(x)}$$

$$Var(\hat{b}) = \frac{E[x^2] * \sigma^2}{m * Var(x)}$$

Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

(4) recentering the data ($x_i = x_i - \mu$)
produce same error on $\hat{w}$
minimize the error on $\hat{b}$ when $\mu = E[x]$

$$Var(\hat{w}) = \frac{\sigma^2}{m * Var(x)}$$

$$= \frac{\sigma^2}{m * \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{m}} = \frac{\sigma^2}{\sum_{i=1}^{m} (x_i - \bar{x})^2}$$

$$= \frac{\sigma^2}{m \sum_{i=1}^{m} \left((x_i - \mu) - \left(\frac{\sum_{i=1}^{m} x_i - \mu}{m} \right) \right)^2}$$

$$= \frac{\sigma^2}{m * Var(x)}$$ → the variance of x doesn't get affected by data shifting

$$Var(\hat{b}) = \frac{E[x^2] * \sigma^2}{m * Var(x)}$$

$$\frac{\frac{1}{m} * \sum_{i=1}^{m} x_i^2 * \sigma^2}{m * \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{m}} \xrightarrow{\text{after shifting}} \frac{\frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)^2 * \sigma^2}{\sum_{i=1}^{m} \left((x_i - \mu) - \frac{\sum_{i=1}^{m} (x_i - \mu)}{m} \right)^2}$$

$$= \frac{Var(x) * \sigma^2}{m * Var(x)} = \frac{\sigma^2}{m}$$

→ the above prove that recentering the data won't change the error on $\hat{w}$ but it does on $\hat{b}$

Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

5)

```
import numpy as np
from sympy.stats import E

def mean(x):
    return sum(x)/len(x)
def var(x):
    m = mean(x)
    all = 0
    for i in x:
        all +=(i-m)**2
    return all/len(x)
def cov(x,y):
    mx = mean(x)
    my = mean(y)
    all = 0
    for i in range(len(x)):
        all += ((x[i]-mx)*(y[i]-my))
    return all/len(x)


m = 200
w = 1
b = 5
sd = 0.1
w_all = []
w_shift = []
b_all = []
b_shift = []
for l in range(1000):
    x = np.random.uniform(100, 102, m)
    y = np.zeros(m)
    e = np.random.normal(0, np.sqrt(sd), m)
    for i in range(len(x)):
        y[i] = (w * x[i]) + b + e[i]
    w_star = cov(x,y)/var(x)
    b_star = mean(y)-(w_star*mean(x))
    w_all.append(w_star)
    b_all.append(b_star)
    x_bar = np.zeros(m)
    for i in range(len(x)):
        x_bar[i] = x[i]-101
    w_star_shift = cov(x_bar,y)/var(x_bar)
    b_star_shift = mean(y)-(w_star_shift*mean(x_bar))
    w_shift.append(w_star_shift)
    b_shift.append(b_star_shift)
print("Mean of w: " + str(mean(w_all)))
print("Variance of w: " + str(var(w_all)))
print("Mean of b: " + str(mean(b_all)))
print("Variance of b: " + str(var(b_all)))
print("Mean of w after data shifting: " + str(mean(w_shift)))
print("Variance of w after data shifting: " + str(var(w_shift)))
print("Mean of b after data shifting: " + str(mean(b_shift)))
print("Variance of b after data shifting: " + str(var(b_shift)))
```

Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

```
Mean of w: 1.0045718203834721
Variance of w: 0.0013528551785233244
Mean of b: 4.5380798086577565
Variance of b: 13.79868224777027
Mean of w after data shifting: 1.0045718203834721
Variance of w after data shifting: 0.0013528551785233248
Mean of b after data shifting: 105.99983366738844
Variance of b after data shifting: 0.0005059696193896366
```

The results make sense and agrees with the above limiting expression, shifting the data doesn't affect w but affect b

Fatima AlSaadeh (fya7)
CS 536 : Regression and Error

6) The linear model is trying to minimize the training error, when the data is shifted the model still gives a straight line shifted on the x-axis for all the points, the y-axis is still the same and the positions of each point to each other and so no change happened on the slope of the linear model

7) $\Sigma = X^T X$

$$\Sigma \to m \begin{bmatrix} 1 & E[x] \\ E[x] & E[x^2] \end{bmatrix}$$

* assume the first dimension in X is 1

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \qquad \Sigma = X^T X = \underset{2*m}{\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_m \end{bmatrix}} \underset{m \times 2}{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}}$$

$$= \begin{bmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{bmatrix}$$ → taking m out of the matrix.

$$\Sigma = m \begin{bmatrix} 1 & E[x] \\ E[x] & E[x^2] \end{bmatrix}$$ and this is how we got that limit.

→ to show that when recentering the data $K(\Sigma')$ is minimized taking $\mu = E[x]$

$$K(\Sigma') = \frac{\text{largest eigen value}}{\text{smallest eigen value}}.$$

before recentering eigen values:

$$\lambda_1 = \frac{(1 + E[x^2]) - \sqrt{(1 + E[x^2])^2 - 4(E[x^2] - E[x]^2)}}{2}$$

$$\lambda_1 = (1 + E[x^2]) - \sqrt{(1 + E[x^2])^2 - 4\,Var(x)} \times \frac{1}{2}$$

$$\lambda_2 = (1 + E[x^2]) + \sqrt{(1 + E[x^2])^2 - 4\,var(x)} \times \frac{1}{2}$$

after recentering :

$$\lambda_1 = \frac{1}{2} * (1+E[x'^2]) - \sqrt{(1+E[x'^2])^2 - 4\,var(x')}$$

$$\lambda_2 = \frac{1}{2} * (1+E[x'^2]) + \sqrt{(1+E[x'^2])^2 - 4\,var(x')}$$

analyzing this for $x' = x - \mu$

$$\lambda_1 = \frac{1}{2} * (1+E[(x-\mu)^2]) - \sqrt{(1+E[(x-\mu)^2])^2 - 4\,var(x-\mu)}$$

$$\lambda_2 = \frac{1}{2} * (1+E[(x-\mu)^2]) + \sqrt{(1+E[(x-\mu)^2])^2 - 4\,var(x-\mu)}$$

$k = \frac{\lambda_2}{\lambda_1} \Rightarrow$ and with $\mu = E[x]$

$$k = \frac{1+\sqrt{1-\frac{4var(x)}{(1+var(x))^2}}}{1-\sqrt{1-\frac{4var(x)}{(1+var(x))^2}}} = \frac{1}{var(x)}$$

by multiplying the numerator and denomenator by the conjugate the solve.

this prove that the condition number is minimized