Q1 (a.) ~~The~~ We are intersteel in prediction. We use profit, number of employees and industry, which are all real values to predict the salary of CEO. This is a regression problem.

Q1 (b.) This is ~~a~~ a clasification problem. We use the parameter to classify whether the product belongs to success or failure.

Q2. (a.) Determine whether people have diabete. The feature can include people's heights and weights.

Determin whether a beverage is successful or not. The features can include price, flavor and sales.

Determin whether a product is qualified or not. The features can include the functionality, the error rate and the life-span.

(b.) Use to predict a person's credit ~~=~~ line. The features can include annual income, debt and properties owned.

Use to predict the ~~income~~ future income of a saleman. The features ~~is~~ can include history sales, company situation.

Use to predict the rating of a film. The feature can include the genre, the ~~aud~~ targeted audience.

Q3 (a.) The SAT score can be a good feature to consider. The SAT score can reflect student's ability to think critically and the results of SAT is highly accesiable.

(b.) This is a discrete value

(c.) The number of books read by the students, the number of sample test done by student prior to the exam

(d.) I think the linear model is reasonable ~~because~~, and the slope should be positive. Because the more students read, the better they can think. ~~Ad~~ And when students do a lot sample tests, we generally believe they do better in the actual exams.

Q4 (a.) $\bar{x} = \dfrac{0+1+2+3+4}{5} = 2.$

$\bar{y} = \dfrac{0+2+3+8+17}{5} = 6.$

(b)

$$S_x^2 = \frac{\Sigma(x-\bar{x})^2}{4} = \frac{4+1+1+4}{4} = \frac{5}{2}$$

$$S_y^2 = \frac{\Sigma(y-\bar{y})^2}{4} = 46.5$$

$$S_{xy}^2 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{4} = 8$$

(c) $w_1 = \frac{Cov(X,Y)}{Var(X)} = \frac{8}{5/2} = 4$

$w_0 = \bar{y} - \frac{16}{5}\bar{x}$

$= 6 - 4 \times 2$

$= -2$

$y = -\frac{2}{5} + \frac{16}{5}x$

(d.) $y = -\frac{2}{5} + \frac{16}{5} \times 2.5$

$y_p = 7.6$.

(e.) $E_{in} = MSE = \frac{1}{N} RSS = \frac{1}{N} \sum^{N} (\hat{y}_i - y_i)^2$

$\hat{y}_1 = -\frac{2}{5}, \quad \hat{y}_2 = \frac{14}{5}, \quad \hat{y}_3 = 6, \quad \hat{y}_4 = \frac{46}{5}, \quad \hat{y}_5 = \frac{62}{5}$

$\frac{1}{N} \sum^{N} (\hat{y}_i - y_i)^2 = \frac{1}{5} \left[\left(-\frac{2}{5}\right)^2 + \left(\frac{4}{5}\right)^2 + (3)^2 + \left(\frac{6}{5}\right)^2 + \left(\frac{13}{5}\right)^2 \right]$

$= 3.6$

(f) R^2 is ~~the~~ indicates the goodness of fit. It shows how well the model explains the variation

$R^2 = 1 - \frac{RSS}{TSS}$ $\quad TSS = \sum_{i=1}^{N} (y_i - \bar{y})^2 = \frac{1}{5}[(36)+16+9+4+81]$

$= 29.2$

$RSS = 18$

$R^2 = 1 - \frac{18}{29.2} = 0.38$

(9.1) When y_5 is changed into 15, the covariance will decrease because the variation of each data is smaller, therefore w_1 will decrease, so is w_0.

Same reason for the second data set, when y_5 is further decreased, covariance is even smaller. w_1 and w_0 are smaller.

Q5. $$J(w_0, w_1) = \frac{1}{2}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)^2$$

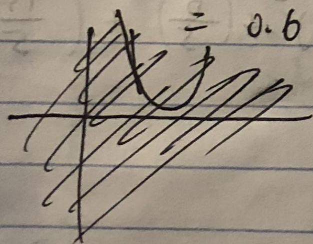
$$\frac{\partial J(w_0, w_1)}{\partial w_0} = \frac{1}{N}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)$$

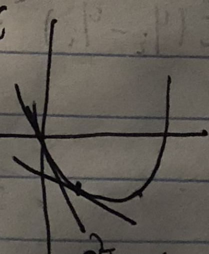
$$\frac{\partial J(w_0, w_1)}{\partial w_1} = \frac{1}{N}\sum_{i=1}^{N}(w_0 + w_1 x_i - y_i)x_i$$

Step 1 $w_{0,1} = 0 - \frac{0.1}{5}(0-2-3-8-17) = 0.6$

Step 2 $w_{0,2} = 0.6 - \frac{0.1}{5}[(0.6-0)+(0.6-2)+(0.6-3)+(0.6-8)+(0.6-17)]$

$= 0.6 + 0.14 = 0.74$





$\frac{\partial J}{\partial w} = -6$, when $w_0 = 0$.

$\frac{\partial J}{\partial w} = -5.4$, when $w_0 = 0.6$.

For w_1.

Step 1 $w_{1,1} = 0 - \frac{0.1}{5}(-2\times1 - 3\times2 - 8\times3 - 17\times4)$

$= 2$

$w_{1,2} = 2 - \frac{0.1}{5}[(2\times1-2)\times1 + (2\times2-3)\times2 + (2\times3-8)\times3 + (2\times4-15)\times4]$

$= 2 + 0.64$

$= 2.64$

Q6. $P(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-\varepsilon_i^2}{2\sigma^2}}$

$P(y|x;w) = \cancel{P()} P(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y_i-(w_0+w_1x_i))^2}{2\sigma^2}}$

The most likely w is the one with the highest probability to have seen the data.

We need to maximize

$L(w) = \prod_{i=1}^{N} P(y_i | \cancel{w} x_i ; w) = \prod_{i=1}^{N} P(\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y_i-(w_0+w_1x_i))^2}{2\sigma^2}}$

$\Rightarrow \log(w) = \log\left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y_i-(w_0+w_1x_i))^2}{2\sigma^2}} \right)$

$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y_i-(w_0+w_1x_i))^2}{2\sigma^2}}$

$= N \cdot \log \frac{1}{\sqrt{2\pi}\sigma} + \frac{1}{2\sigma^2} \times \sum_{i=1}^{N} -(y_i-(w_0+w_1x_i))^2$

$\Rightarrow$ We need to minimize $(y_i - (w_0 + w_1 x_i))^2$

$\frac{1}{2} \sum_{i=1}^{5} (y_i - (w_0 + w_1 x_i))^2$

$w_0 = -1,\ w_1 = 4,$

$\frac{1}{2}[(2-(-1+4\times1))^2 + (3-(-1+4\times2))^2 + (8-(-1+4\times3))^2 + (17-(-1+4\times4))^2]$

$= \cancel{}\ 15$

$w_0 = -2,\ w_1 = 4$

$\frac{1}{2}[[2-(-2+4\times1)]^2 + [3-(-2+8)]^2 + [8-(-2+12)]^2 + [17-(-2+16)]^2]$

$= 11$

$w_0 = -2,\ w_1 = 3.$

$\frac{1}{2}[[2-(-2+3\times1)]^2 + [3-(-2+6)]^2 + [8-(-2+9)]^2 + [17-(-1+12)]^2].$

$= 19.5$

~~[illegible]~~ $w_0 = -2,\ w_1 = 4$ yield the smallest error

Q7. (a.) $Z(t) = Z_0 e^{-\alpha t}$

$\cancel{\log Z(t) = \log Z_0 e^{-\alpha t}}$ $\quad \ln Z(t) = \ln Z_0 e^{-\alpha t}$

$\cancel{\log Z(t) = \log Z_0 - \alpha t}$ $\quad \ln Z(t) = \ln Z_0 - \alpha t$

(b.) $RSS = \sum_{i=1}^{N} (\log(Z) - \log Z_0 + \alpha t)^2$

$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i \qquad \overline{\ln(Z(t))} = \frac{1}{N}\sum_{i=1}^{N} \ln Z(t)$

$Var(t) = \frac{1}{N}\sum(t_i - \bar{t})^2 \qquad Var(\ln(Z)) = \frac{1}{N}\sum(\ln Z(t) - \overline{\ln Z})^2$

$Cov = \frac{1}{N}\sum(t_i - \bar{t})(\ln Z(t) - \overline{\ln Z})$

$\therefore -\alpha = \frac{\sum(t_i - \bar{t})(\ln Z - \overline{\ln Z})}{\sum(t_i - \bar{t})^2} \qquad \ln Z_0 = \overline{\ln Z(t)} - \frac{\sum(t_i - \bar{t})(\ln Z - \overline{\ln Z})}{\sum(t_i - \bar{t})^2} \times \bar{t}$

Q8. (a.) $RSS = \sum_{i=1}^{N} (y_i - w x_i)^2$

(b.) $\frac{\partial RSS}{\partial w} = 2\sum_{i=1}^{N} (y_i - w x_i) \times x_i = 0.$

$\sum_{i=1}^{N} x_i y_i - w x_i^2 = 0.$

$\sum_{i=1}^{N} x_i y_i = \cancel{N} w \sum_{i=1}^{N} x_i^2$

$w = \frac{\sum x_i y_i}{\sum x_i^2}$

In [1]:
```
import numpy as np
```

In [ ]:
```
def fit_linear(t,z0):
    #need to fit the condition in the homework
    x = t
    y = np.log(z0)

    x_mean = x.mean()
    y_mean = y.mean()

    # TODO - Find sxx, sxy and syy - 5 points
    sxx = np.var(x)
    syy = np.var(y)
    sxy = np.cov(x,y)[0][1]

    # TODO - Find w0 and w1 - 5 points
    w1 = sxy/sxx
    w0 = y_mean - w1 * x_mean

    #again fit the condition in the homework
    w1 = -w1
    w0 = np.exp(w0)
    return w0, w1
```