Regression and Error

1. a linear model that minimizes the training error

Let $I = \sum_{i=1}^{m}(wx_i + b - y_i)^2$

To compute w:

$\frac{dI}{dw} = 2\sum_{i=1}^{m}(wx_i + b - y_i)x_i$

Let $\frac{dI}{dw} = 0$:

$\hat{w} = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2}$, where $\bar{y} = \frac{\sum_{i=1}^{m} y_i}{m}$, $\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$, $\bar{yx} = \frac{\sum_{i=1}^{m} x_i y_i}{m}$, $\bar{x^2} = \frac{\sum_{i=1}^{m} x_i^2}{m}$

To compute $\hat{b}$, we have:

$\frac{dI}{db} = 2\sum_{i=1}^{m}(wx_i + b - y_i)$

Let $\frac{dI}{db} = 0$:

$\hat{b} = \bar{y} - \hat{w}\bar{x}$

2. εi ~ N(0,σ2)

A. mean

$y_i = wx_i + b + \epsilon_i \Rightarrow y_i \sim N(wx_i + b, \sigma^2)$.

$$E(\hat{w}) = E\left(\frac{m\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m\sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2}\right) = \frac{m\sum_{i=1}^{m} x_i E(y_i) - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} E(y_i)}{m\sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2} = \frac{m\sum_{i=1}^{m} x_i (wx_i + b) - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} (wx_i + b)}{m\sum_{i=1}^{m} x_i^2 - (\sum_{i=1}^{m} x_i)^2} = w$$

$$E(\hat{b}) = E(\bar{y}) - E(\hat{w})\bar{x} = w\bar{x} + b - w\bar{x} = b$$

B. variance

Notice that $w = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x^2} - \bar{x}^2} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$

$$var(\hat{w}) = var\left(\frac{\sum_{i=1}^{m}(x_i - \bar{x})y_i}{\sum_{i=1}^{m}(x_i - \bar{x})^2} - \frac{\sum_{i=1}^{m}(x_i - \bar{x})\bar{y}}{\sum_{i=1}^{m}(x_i - \bar{x})^2}\right) = \left(\frac{\sum_{i=1}^{m}(x_i - \bar{x})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}\right)^2\sigma^2 - 0 = \frac{\sigma^2}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

$$var(\hat{b}) = var(\bar{y} - \hat{w}\bar{x}) = var(\bar{y}) + \bar{x}^2 var(\hat{w}) = \frac{\sigma^2}{n} + \frac{\bar{x}^2\sigma^2}{\sum_{i=1}^{m}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{m} x_i^2\sigma^2}{m\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

3. expectation E(x) and variance Var(x)

$$E(x) \approx \bar{x}$$

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

$$var(\hat{w}) \approx \frac{\sigma^2}{mvar(x)}$$

$$var(\hat{b}) \approx \frac{\sigma^2}{mvar(x)}\frac{\sum_{i=1}^{m} x_i^2}{m} = \frac{\sigma^2}{mvar(x)}E(x^2)$$

4. recentering the data

Notice that after recentering the data,

$var(x') = var(x),\ E(x') = 0$

Therefore, $E({x'}^2) = var(x') + E(x')^2 = \min_x var(x) + E(x)^2 = \min_x E(x^2)$

Hence, $var(\hat{w}') = var(\hat{w}),\ var(\hat{b}') = \min_x var(\hat{b})$

5. Verify

Here is the result: (Fig.1)

```
D:\Users\endlesstory\Desktop\536\hw5>python RP.py
E(w):
practice: 0.99890, theory: 1.00000
E(w'):
practice: 0.99890, theory: 1.00000
E(b):
practice: 5.11206, theory: 5.00000
E(b'):
practice: 106.00088, theory: 106.00000
var(w):
practice: 0.00151, theory: 0.00150
var(w'):
practice: 0.00151, theory: 0.00150
var(b):
practice: 15.39615, theory: 15.30200
var(b'):
practice: 0.00051, theory: 0.00050
```

Fig.1 the expected value and variance of $\hat{w},\ \hat{w'},\ \hat{b},\ \hat{b'}$.

Notice that practice values are closed to the theory values. Therefore it agrees with the above expressions.

6. no change in the estimate of the slope

Notice that when we recenter the data, all what we have done is nothing but move the whole graph along the x-axis for some distance. The estimated line is also moved in the same way, and therefore, the slope will not be any different.

7. augmenting the data

A. XTX

Let $X = [1_{m\times 1}, x_{m\times 1}]$, where $x_{m\times 1} = (x_1, ..., x_m)^T$

$$X^T X = \begin{bmatrix} 1^T 1 & 1^T x \\ x^T 1 & x^T x \end{bmatrix}$$

Notice that $1^T 1 = m,\ 1^T x = x^T 1 = \sum_{i=1}^{m} x_i,\ x^T x = \sum_{i=1}^{m} x_i^2$

Therefore, $\Sigma \approx m \begin{bmatrix} 1 & E(x) \\ E(x) & E(x^2) \end{bmatrix}$

B. recentering the data

Let $\det(\lambda I - \Sigma) = 0$:
$(\frac{\lambda}{m})^2 - (1 + E(x^2))\frac{\lambda}{m} + E(x^2) - E(x)^2 = \frac{\lambda}{m}^2 - (1 + E(x^2))\frac{\lambda}{m} + var(x) = 0$

Therefore $\lambda = \frac{m}{2}((1 + E(x^2)) \pm \sqrt{(1 + E(x^2))^2 - 4var(x)}$

Therefore, $\kappa(\Sigma) = \frac{1+var(x)+E(x)^2+\sqrt{(1+var(x)+E(x)^2)^2-4var(x)}}{1+var(x)+E(x)^2-\sqrt{(1+var(x)+E(x)^2)^2-4var(x)}}$

Notice that when $E(x)^2 = 0$, namely $E(x) = 0$, we have the minimum $\kappa(\Sigma) = \frac{1}{var(x)}$.
In this case, $x' = x - \mu$, where $\mu = E(x)$.