# Practical Class 03

## Inner Product and Linear Regression

1. Suppose that $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are three vectors in $\mathbb{R}^3$ such that

$$2\mathbf{x} + 2\mathbf{y} + \mathbf{z} = \mathbf{0}$$

and $|\mathbf{x}| = 3$, $|\mathbf{y}| = 4$, $|\mathbf{z}| = 7$. Use inner product (dot product) to find the angle between $\mathbf{x}$ and $\mathbf{y}$.

**Solution:** Let

$$\mathbf{x} = (x_1, x_2, x_3),\ \ \mathbf{y} = (y_1, y_2, y_3) \text{ and } \mathbf{z} = (z_1, z_2, z_3).$$

Then the condition $2\mathbf{x} + 2\mathbf{y} + \mathbf{z} = 0$ can be written as

$$2x_1 + 2y_1 + z_1 = 0,\ \ [2x_2 + 2y_2 + z_2 = 0,\ \ [2x_3 + 2y_3 + z_3 = 0,$$

or equivalently,

$$2x_1 + 2y_1 = -z_1,\ \ [2x_2 + 2y_2 = -z_2,\ \ [2x_3 + 2y_3 = -z_3.$$

Recall that the inner(dot) product of $\mathbf{x}$ and $\mathbf{y}$ is defined to be $x_1y_1 + x_2y_2 + x_3y_3$, take squares of each of the above equation yields

$$4x_1^2 + 4y_1^2 + 8x_1y_1 = z_1^2,\ \ 4x_2^2 + 4y_2^2 + 8x_2y_2 = z_2^2,\ \ 4x_3^2 + 4y_3^2 + 8x_3y_3 = z_3^2.$$

Sum up these three equations and we get

$$4(x_1^2 + x_2^2 + x_3^2) + 4(y_1^2 + y_2^2 + y_3^2) + 8(x_1y_1 + x_2y_2 + x_3y_3) = z_1^2 + z_2^2 + z_3^2,$$

that is

$$4|\mathbf{x}|^2 + 4|\mathbf{y}|^2 + 8 < \mathbf{x}, \mathbf{y} >= |\mathbf{z}|^2.$$

Substitute the value in, we get

$$< \mathbf{x}, \mathbf{y} >= -6.375 = |\mathbf{x}||\mathbf{y}|cos(\theta),$$

where $\theta$ is angle between vector $\mathbf{x}$ and $\mathbf{y}$. Therefore,

$$cos(\theta) = \frac{< \mathbf{x}, \mathbf{y} >}{|\mathbf{x}||\mathbf{y}|} = \frac{-6.375}{3 \times 4} = -0.53125,$$

and

$$\theta = \ 2.13.$$

2. As seen in the lecture slides, the parameter vector for a general linear regression with a bias term

$$h_\theta(\mathbf{x}) = \theta[0] + \theta[1]x[1] + \theta[2]x[2] + \cdots + \theta[n]x[n]$$

can be found in the matrix form by

$$\theta = (XX^T)^{-1}X\mathbf{y},$$

where

$$X = \begin{bmatrix} x_1[1] & x_2[1] & \cdots & x_m[1] \\ x_1[2] & x_2[2] & \cdots & x_m[2] \\ \vdots & \vdots & \ddots & \vdots \\ x_1[n] & x_2[n] & \cdots & x_m[n] \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Now, consider a simple model with $n = 1$, that is,

$$y = \theta[0] + \theta[1]x.$$

Try to deduce the general formula for parameter $\theta[0]$ (bias) and $\theta[1]$ using matrix multiplication for a data set with m training samples.

**Solution:** Let's denote the traing data as $\{(x_i, y_i\}_{i=1}^m$. The augmented data matrix in this case is

$$X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix},$$

and

$$XX^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \times \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} = \begin{bmatrix} m & m\bar{x} \\ m\bar{x} & \sum_{i=1}^m x_i^2 \end{bmatrix},$$

noticing that $\sum_{i=1}^m x_i = m\bar{x}$. Recall that the inverse for $2 \times 2$ matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Thus, we have

$$\begin{aligned}
\theta = \begin{bmatrix} \theta[0] \\ \theta[1] \end{bmatrix} &= \frac{1}{m\sum_{i=1}^m x_i^2 - (m\bar{x})^2} \begin{bmatrix} \sum_{i=1}^m x_i^2 & -m\bar{x} \\ -m\bar{x} & m \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \\
&= \frac{1}{m\sum_{i=1}^m x_i^2 - (m\bar{x})^2} \begin{bmatrix} \sum_{i=1}^m x_i^2 - m\bar{x}x_1 & \cdots & \sum_{i=1}^m x_i^2 - m\bar{x}x_m \\ -m\bar{x} + mx_1 & \cdots & -m\bar{x} + mx_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \\
&= \frac{1}{m\sum_{i=1}^m x_i^2 - (m\bar{x})^2} \begin{bmatrix} m\bar{y}\left(\sum_{i=1}^m x_i^2\right) - m\bar{x}\left(\sum_{i=1}^m x_i y_i\right) \\ m\sum_{i=1}^m x_i y_i - m^2\bar{x}\bar{y} \end{bmatrix},
\end{aligned}$$

that is

$$\theta[0] = \frac{\bar{y}\left(\sum_{i=1}^m x_i^2\right) - \bar{x}\left(\sum_{i=1}^m x_i y_i\right)}{\sum_{i=1}^m x_i^2 - m\bar{x}^2} \quad \text{and} \quad \theta[1] = \frac{\sum_{i=1}^m x_i y_i - m\bar{x}\bar{y}}{\sum_{i=1}^m x_i^2 - m\bar{x}^2}$$

3. A professor in the faculty of Engineering polled a dozen colleagues ($m$) about the number of professional meetings they attended in the past five years ($x$) and the number of papers they submitted to refereed journals ($y$) during the same period. The summary data are given as follows:

$$m = 12,\ \bar{x} = 4,\ \bar{y} = 12,$$

$$\sum_{i=1}^{12} x_i^2 = 232,\ \sum_{i=1}^{12} x_i y_i = 318.$$

Fit a simple linear regression model between x and y by finding the estimate of the intercept and slope using your results from the previous question. Comment on whether attending more professional meetings would result in publishing more papers.

**Solution:** From the previous question

$$\theta[0] = \frac{\bar{y}\left(\sum_{i=1}^{m} x_i^2\right) - \bar{x}\left(\sum_{i=1}^{m} x_i y_i\right)}{\sum_{i=1}^{m} x_i^2 - m\bar{x}^2} = \frac{(12)(232) - (4)(318)}{232 - (12)(4^2)} = \frac{189}{5} = 37.8,$$

and

$$\theta[1] = \frac{\sum_{i=1}^{m} x_i y_i - m\bar{x}\bar{y}}{\sum_{i=1}^{m} x_i^2 - m\bar{x}^2} = \frac{318 - (12)(4)(12)}{232 - (12)(4^2)} = -\frac{129}{20} = -6.45.$$

So the simple linear regression line is

$$y = 37.8 - 6.45x.$$

It appears that attending professional meetings would not result in publishing more papers (due to the negative slope).

4. A study was made on the amount of converted sugar in a certain process at various temperatures. The data was recorded as follows:

| **Temperature (x)** | 1.0 | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Converted sugar (y)** | 8.1 | 7.8 | 8.5 | 9.8 | 9.5 | 8.9 | 8.6 | 10.2 | 9.3 | 9.2 | 10.5 |

(a) Fit a linear regression line to the data using your result to Q2 and plot the regression line on a graph with the data.

(b) Estimate the mean amount of converted sugar produced when the temperature is 1.75.

(c) Plot the residuals versus temperature. Comment.

**Solution:**

(a) With the data set, we have

$$n = 11,\ \sum_{i=1}^{n} x_i = 16.5,\ \sum_{i=1}^{n} y_i = 100.4,\ \sum_{i=1}^{n} x_i^2 = 25.85,\ \sum_{i=1}^{n} x_i y_i = 152.59.$$

Using previous results,

$$\theta[0] = \frac{(100.4/11)(25.85) - (16.5/11)(152.59)}{25.85 - 16.5^2/11} = 6.4132,$$

$$\theta[1] = \frac{152.59 - (16.5)(100.4)/11}{25.85 - 16.5^2/11} = 1.8091.$$

Hence

$$\hat{y} = 6.4136 + 1,8091x.$$

(b) For $x = 1.75$, $\hat{y} = 6.4136 + 1.8091(1.75) = 9.580$.

(c) See Figure below.Residuals appear to be random as desired.