

Practical Class 03

Inner Product and Linear Regression

- 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} .

- 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

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

where

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

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.

- 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.

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.