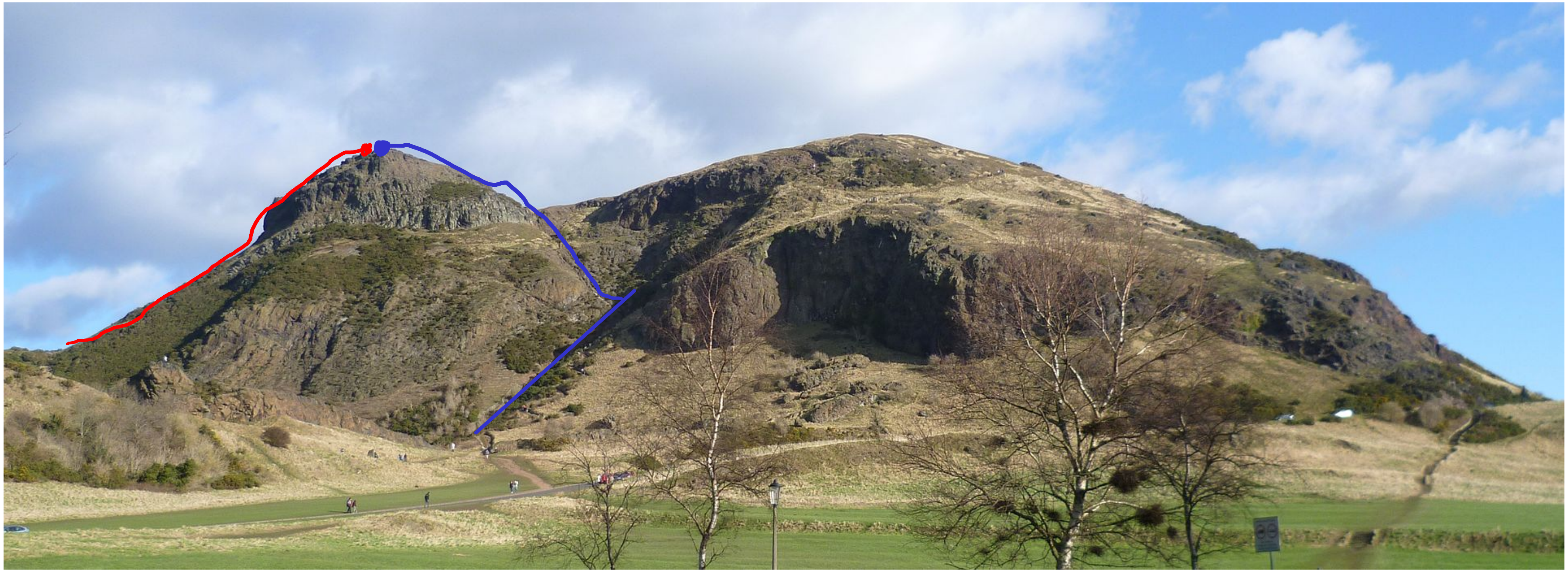
Foundations of Data Science: Multiple regression - Derivation of coefficients

Reading: Modern mathematical statistics with applications, pp. 705-707



Kim T r ayno

Part 1 = simplifying

Part 2 = Coefficients themselves

Next video = properties of equation for coeffs.

- Linear algebra
- Calculus
- Definitions

Least squares

Generally : k indep. variables, 2 for now

$$f(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}))^2$$

$$\frac{\partial f}{\partial \beta_0} = \sum_{i=1}^{n} (-2)(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2}) = 0$$

1. Divide by $-2n$ 2. $\bar{x}^{(j)} = \frac{1}{n} \sum_i x_{ij}$

$$\bar{y} - \beta_0 - \beta_1 \bar{x}^{(1)} - \beta_2 \bar{x}^{(2)} = 0$$

$$\Rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x}^{(1)} - \beta_2 \bar{x}^{(2)}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}^{(1)} - \beta_2 \bar{x}^{(2)} \quad ①$$

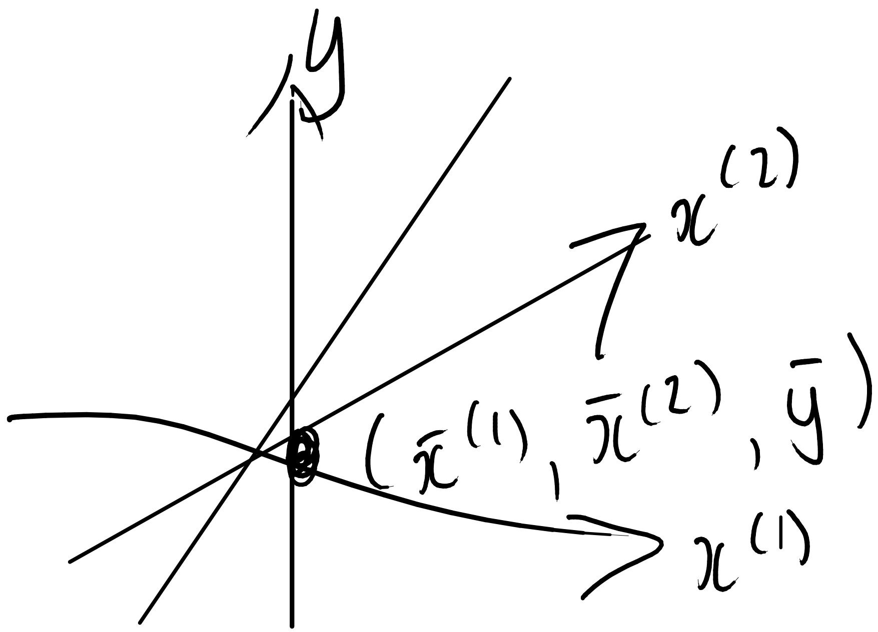
$$y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} \quad ②$$

$$y = \bar{y} - \beta_1 \bar{x}^{(1)} - \beta_2 \bar{x}^{(2)} + \beta_1 x^{(1)} + \beta_2 x^{(2)}$$

$$y - \bar{y} = \beta_1 \left(x^{(1)} - \bar{x}^{(1)} \right) + \beta_2 \left(x^{(2)} - \bar{x}^{(2)} \right)$$

$$y_i^* = y_i - \bar{y}$$

$$x_{ij}^* = x_{ij} - \bar{x}^{(j)}$$



Derivation: Part 2

$$f^*(\beta_1, \beta_2) = \sum_i (y_i^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^*)^2$$

$$\frac{\partial f^*}{\partial \beta_1} = \sum_i (-2x_{i1}^*)(y_i^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^*) = 0$$

$$\frac{\partial f^*}{\partial \beta_2} = \sum_i (-2x_{i2}^*)(y_i^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^*) = 0$$

$$\begin{pmatrix} \sum_i x_{i1}^* (y_i^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^*) \\ \sum_i x_{i2}^* (y_i^* - \beta_1 x_{i1}^* - \beta_2 x_{i2}^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \sum_i x_{i1}^* (\beta_1 x_{i1}^* + \beta_2 x_{i2}^*) \\ \sum_i x_{i2}^* (\beta_1 x_{i1}^* + \beta_2 x_{i2}^*) \end{pmatrix} = \begin{pmatrix} \sum_i x_{i1}^* y_i^* \\ \sum_i x_{i2}^* y_i^* \end{pmatrix}$$

$$\begin{pmatrix} \sum_i x_{i1}^* y_i^* \\ \sum_i x_{i2}^* y_i^* \end{pmatrix} = X^\top \bar{y}$$ ← moment matrix

$$X = \begin{array}{cc} \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix} & n \\ 2 & \end{array}$$

Design matrix

Regressor matrix

$$\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\begin{pmatrix} \sum_i x_{i1}^* (\beta_1 x_{i1}^* + \beta_2 x_{i2}^*) \\ \sum_i x_{i2}^* (\beta_1 x_{i1}^* + \beta_2 x_{i2}^*) \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} \sum_i x_{i1}^* x_{i1}^* & \sum_i x_{i1}^* x_{i2}^* \\ \sum_i x_{i2}^* x_{i1}^* & \sum x_{i2}^* x_{i2}^* \end{pmatrix}}_{X^T X} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

1. $(n-1)\, s_2^2$ — Diagonals   sample variance $x^{(1)}$ $x^{(2)}$

   $(n-1)\, s_{12}$ — off diagnals — sample covariance $x^{(1)}$ $x^{(2)}$

2. The 2x 2matrix can be written $X^T X$
   the normal matrix

3 Covariance matrix $S = \frac{1}{n-1} X^T X$

$$\begin{pmatrix} \sum_i x_{i1}^* (\beta_1 x_{i1}^* + \beta_2 x_{i2}^*) \\ \sum_i x_{i2}^* (\beta_1 x_{i1}^* + \beta_2 x_{i2}^*) \end{pmatrix} = \begin{pmatrix} \sum_i x_{i1}^* y_i^* \\ \sum_i x_{i2}^* y_i^* \end{pmatrix}$$

$$X^T X \underline{\hat{\beta}} = X^T \underline{y}$$

$$\Rightarrow \underline{\underline{\underline{\hat{\beta}} = (X^T X)^{-1} X^T \underline{y}}}$$

$$\underline{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}^{(1)} - \hat{\beta}_2 \bar{x}^{(2)}$$