Part 1.

①

$$\nabla f(x,y) = \begin{bmatrix} \frac{df(x,y)}{dx} \\ \frac{df(x,y)}{dy} \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

②

$$\nabla f(x) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$$

③ $f_x(x,y) = 2A(x - x_0)$

$f_y(x,y) = 2B(y - y_0)$

④ $x^T = (3\ 1\ 4)$ $[1 \times 3]$

$y^T = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}$ $[3 \times 1]$

$B^T = \begin{pmatrix} 3 & 5 & 1 \\ 5 & 2 & 4 \end{pmatrix}$ $[2 \times 3]$

$x \cdot x$, $x \cdot y^T$ not defined

$$x \times y = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} (2\ 5\ 1) = \begin{pmatrix} 6 & 15 & 3 \\ 2 & 5 & 1 \\ 8 & 20 & 4 \end{pmatrix}$$

$$y \times x = (2\ 5\ 1) \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} = 2 \times 3 + 5 \times 1 + 1 \times 4 = 15$$

$$A \times x = \begin{pmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 12+5+8 \\ 9+1+20 \\ 18+4+12 \end{pmatrix} = \begin{pmatrix} 25 \\ 30 \\ 34 \end{pmatrix}$$

$A \times B =$

B.reshape(1,6) = (3 5 5 2 1 4)

Linear Least squares (LLS):

① Loss surface: $L(P) = L(m, b) = \sum_{i=1}^{N} (\hat{y}_i - M(\hat{x}_i, m, b))^2$

$\min L(P)$.

$$\frac{dL(m,b)}{dm} = \frac{d \sum_{i=1}^{N} (\hat{y}_i - (m\hat{x}_i + b))^2}{dm}$$

$$= \sum_{i=1}^{N} (-\hat{x}_i) \cdot 2 (\hat{y}_i - (m\hat{x}_i + b))$$

$$= 2 \sum_{i=1}^{N} \hat{x}_i (m\hat{x}_i + b - \hat{y}_i) = 0$$

$$m \sum_{i=1}^{N} \hat{x}_i^2 + b \sum_{i=1}^{N} \hat{x}_i - \sum_{i=1}^{N} \hat{x}_i \hat{y}_i = 0$$

$$m = \frac{\sum_{i=1}^{N} \hat{x}_i \hat{y}_i - b \sum_{i=1}^{N} \hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i^2}$$

$$\frac{dL(m,b)}{db} = \frac{d \sum_{i=1}^{N} (\hat{y}_i - (m\hat{x}_i + b))^2}{db}$$

$$= \sum_{i=1}^{N} (-1) \cdot 2 (\hat{y}_i - (m\hat{x}_i + b))$$

$$= 2 \sum_{i=1}^{N} (m\hat{x}_i + b - \hat{y}_i) = 0$$

$$m \sum_{i=1}^{N} \hat{x}_i + Nb - \sum_{i=1}^{N} \hat{y}_i = 0$$

$$b = \frac{\sum_{i=1}^{N} \hat{y}_i}{N} - m \frac{\sum_{i=1}^{N} \hat{x}_i}{N}$$

$$= \bar{y} - m\bar{x}$$

$$m = \frac{\sum_{i=1}^{N} \hat{x}_i \hat{y}_i - b \sum_{i=1}^{N} \hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i^2} = \frac{\sum_{i=1}^{N} \hat{x}_i \hat{y}_i - \bar{y} \sum_{i=1}^{N} \hat{x}_i + m\bar{x} \sum_{i=1}^{N} \hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i^2}$$

$$m = \frac{\sum_{i=1}^{N} \hat{x}_i \hat{y}_i - \bar{y} \sum_{i=1}^{N} \hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i^2 - \bar{x} \sum_{i=1}^{N} \hat{x}_i}$$

$$m = \frac{\sum_{i=1}^{N} \hat{x}_i \hat{y}_i - \bar{y} \sum_{i=1}^{N} \hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i^2 - \bar{x} \sum_{i=1}^{N} \hat{x}_i}$$

$$= \frac{\frac{N \sum_{i=1}^{N} \hat{x}_i \hat{y}_i}{N} - \frac{\sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{y}_i}{N}}{\frac{N \sum_{i=1}^{N} \hat{x}_i^2}{N} - \frac{\left(\sum_{i=1}^{N} \hat{x}_i\right)^2}{N}}$$

$$= \frac{N \sum_{i=1}^{N} \hat{x}_i \hat{y}_i - \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{y}_i}{N \sum_{i=1}^{N} \hat{x}_i^2 - \left(\sum_{i=1}^{N} \hat{x}_i\right)^2}$$

$$\operatorname{cov}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - \bar{x})(\hat{y}_i - \bar{y})$$

$$\operatorname{Var}(x) = \frac{1}{N} \sum_{i=1}^{N} (\hat{x}_i - \bar{x})^2$$

$$\frac{\operatorname{cov}(x, y)}{\operatorname{Var}(x)} = \frac{\sum_{i=1}^{N} (\hat{x}_i - \bar{x})(\hat{y}_i - \bar{y})}{\sum_{i=1}^{N} (\hat{x}_i - \bar{x})^2} = \frac{\sum_{i=1}^{N} (\hat{x}_i \hat{y}_i) - \sum_{i=1}^{N} \hat{x}_i \bar{y} - \sum_{i=1}^{N} \hat{y}_i \bar{x} + \sum_{i=1}^{N} \bar{x}\bar{y}}{\sum_{i=1}^{N} \hat{x}_i^2 - 2\sum_{i=1}^{N} \hat{x}_i \bar{x} + \sum_{i=1}^{N} \bar{x}^2}$$

$$= \frac{\frac{N \sum_{i=1}^{N} (\hat{x}_i \hat{y}_i)}{N} - \frac{2 \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{y}_i}{N} + \frac{N \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{y}_i}{N \cdot N}}{\frac{N \sum_{i=1}^{N} \hat{x}_i^2}{N} - \frac{2 \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{x}_i}{N} + \frac{N \cdot \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{x}_i}{N \cdot N}}$$

$$= \frac{N \sum_{i=1}^{N} (\hat{x}_i \hat{y}_i) - \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{y}_i}{N \sum_{i=1}^{N} \hat{x}_i^2 - \sum_{i=1}^{N} \hat{x}_i \sum_{i=1}^{N} \hat{x}_i} = m$$

$$\therefore m = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}$$

since $b = \bar{y} - m\bar{x}$ $\therefore b = \bar{y} - \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} \bar{x}$

Linear least squares (LLS): Multi-variable (Extra credit).

$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ $\quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & N_N \end{bmatrix}$ $\quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$

$f(x_i) = x^{(i)}\vec{\beta}$

loss function

$$L(\beta) = \sum_{i=1}^{N} (f(x_i) - y_i)^2$$

$$= (X\vec{\beta} - \vec{y})^T (X\vec{\theta} - \vec{y})$$

$$\underline{\nabla L(\beta)} = \nabla_\beta (\vec{x}\vec{\beta} - \vec{y})^T (\vec{x}\vec{\beta} - \vec{y})$$

$$= \nabla_\beta (\beta^T X^T - y^T)(X\beta - y)$$

$$= \nabla_\beta (\beta^T X^T X \beta - \beta^T X^T y - y^T X \beta + y^T y)$$

$$= X^T X \beta + X^T X \beta - X^T y - X^T y$$

$$= X^T X \theta - X^T y = \vec{0}$$

$$\beta = (X^T X)^{-1} X^T y.$$