1. Compute the gradient vector for a plane in 3-D space.

$$z = f(x,y) = ax + by + c$$

$$\nabla(f) = \begin{bmatrix} \frac{\partial}{\partial x} ax+by+c \\ \frac{\partial}{\partial y} ax+by+c \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

2. Compute the gradient vector for hyperplane.

$$z = f(x_1, x_2, \cdots, x_N) = \sum_{i=1}^{N} a_i(x_i - b_i) = a_1x_1 + a_2x_2 + \cdots + a_Nx_N + d$$

$$\nabla(z) = \begin{bmatrix} \frac{\partial}{\partial x_1} a_1x_1 + a_2x_2 + \cdots + a_Nx_N + d \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_N \end{bmatrix}$$

3. Compute the partial derivative of the paraboloid function.

$$z = f(x,y) = A(x-x_0)^2 + B(y-y_0)^2 + c$$

$$f_x(x,y) = \frac{\partial z}{\partial x} = \frac{\partial A(x-x_0)^2}{\partial x}$$

$$= 2A(x-x_0)\cdot 2A$$

$$= 4A^2(x-x_0)$$

$$f_y(x,y) = 4B^2(y-y_0)$$

4. Given the following matrices and vectors:

$$x = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} \quad y = \begin{pmatrix} 2 & 5 & 1 \end{pmatrix}$$

$$A = \begin{bmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 5 \\ 5 & 2 \\ 1 & 4 \end{bmatrix}$$

Compute the following qualities and specify the shape of the output, if an operation is not defined then just say not defined.

$x^T = [3 \quad 1 \quad 4]$ $\quad y^T = [$

$\underline{1 \times 3}$ $\quad 3 \times$

$B^T = \begin{bmatrix} 3 & 5 & 1 \\ 5 & 2 & 4 \end{bmatrix}$ $\quad x \cdot x = 3 \quad + 4^2 = 26$

$\underline{2 \times 3}$ $\quad$ $\underline{\text{scalar}}$

$x \cdot y^T = 6 + 5 + 4 = 15$

$\underline{\text{scalar}}$

$x \times y = [-19 \quad 5 \quad 13]$

$\underline{1 \times 3}$

$y \times x = [19 \quad -5 \quad -13]$

$\underline{1 \times 3}$

$A \times x = \begin{bmatrix} 25 \\ 30 \\ 34 \end{bmatrix}$

$\underline{3 \times 1}$

$A \times B = \begin{bmatrix} 39 & 38 \\ 19 & 37 \\ 41 & 50 \end{bmatrix}$

$\underline{3 \times 2}$

$$B.\text{reshape}(1, 6) = [3 \quad 5 \quad 5 \quad 2 \quad 1 \quad 4]$$

$\underline{1 \times 6}$

5. Linear least squares (LLS): Single-variable.
    - Use calculus to analytically derive the expression for single variable linear regression fitting parameters using the sum of square error as the loss function (show your work).

$$y = M(x|P) = mx + b$$

$$p = (p_0, p_1) = (m, b)$$

$$\mathcal{L}(p) = \mathcal{L}(m, b) = \sum_i^N (\hat{y}_i - M(\hat{x}_p, m, b))^2$$

$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - (\hat{m}x_i + \hat{b}))^2$$

$$\frac{\partial}{\partial \hat{b}} SSE = \sum \frac{\partial}{\partial \beta_0} (y_i - (\hat{b} + \hat{m}x_i))^2$$

$$= \sum 2 (y_i - (\hat{m}x_i + \hat{b})) \cdot (-1)$$

$$= -2 \sum (Y_i - (\hat{m}x_i + \hat{b}_1))$$

$$\frac{\partial}{\partial \hat{\beta}_1} \sum (y_i - (\hat{m}x_i + b))^2 = \sum \frac{\partial}{\partial \hat{\beta}_1} (y_i - (\hat{m}x_i + b))^2$$

$$= \sum 2 (y_i - (\hat{m}x + \hat{b}))(-x_i)$$

$$= -2 \sum x_i (y_i - (\hat{m}x + b))$$

$$-\cancel{2} \sum (y_i - (\hat{m}x_i + \hat{b})) = 0$$

$$-\cancel{2} \sum x_i (y_i - (\hat{m}x_i + \hat{b})) = 0$$

$$\sum (y_i - (\hat{m}x_i + \hat{b})) = 0 \Rightarrow \sum y_i - \sum \hat{b} - \sum \hat{m}x_i = 0$$

$$\boxed{\hat{b} = \bar{y} - \hat{m}\bar{x}}$$

$$\sum x_i(Y_i - (\hat{m}x + \hat{b})) = 0$$

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

$$\sum x_i(y_i - (\bar{y} - \hat{m}\bar{x} + \hat{m}x_i)) = 0$$

$$\sum x_i(y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})) = 0$$

$$\sum x_i(y_i - \hat{y} - \hat{m}(x_i - \hat{x})) = 0$$

$$\sum x_i(y_i - \bar{y}) - \sum \hat{m} x_i(x_i - \bar{x}) = 0$$

$$\sum x_i(y_i - \bar{y}) = \hat{m} \sum x_i(x_i - \bar{x})$$

$$\hat{m} = \frac{\sum x_i(Y_i - \hat{Y})}{\sum x_i(x_i - x)} = \frac{cov(x,y)}{Var(x)}$$

$$\boxed{\hat{m} = \frac{cov(x,y)}{Var(x)}}$$

□ Q.E.D.

6. Linear least squares (LLS): multi-variable.
    - Matrix calculus to analytically derives expression for two variable linear regression

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

$$Y_1 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{12} + \varepsilon_1$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 x_n + \beta_2 x_{n2} + \varepsilon_i$$

$$\Rightarrow \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 x_1 + \beta_2 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n + \beta_2 x_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon$$

$$\varepsilon = Y - X\beta$$

$$\sum \varepsilon_i^2 = [\varepsilon_1 \ \varepsilon_2 \cdots \varepsilon_n] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon'\varepsilon$$

$$\min(\varepsilon'\varepsilon) \Rightarrow \min (Y - X\beta)'(Y - X\beta)$$

$$\frac{\partial}{\partial\beta} \varepsilon'\varepsilon = \frac{\partial}{\partial\beta}((Y-X\beta)'(Y-X\beta)) = -2X'(Y-X\beta)$$

$$\frac{\partial}{\partial\beta}\varepsilon'\varepsilon = 0 \Rightarrow -2X'(Y-X\beta) = 0$$

$$\Rightarrow X'Y = X'X\beta \Rightarrow \beta = \frac{X'Y}{X'X}$$

$$\vec{w} = (X'X)^{-1}X'Y$$

$\square$ Q.E.D.