- Compute the gradient vector for a plane in 3D space (0.5 point)

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

$f_x(x,y) = \frac{\partial f}{\partial x} = a$ $\quad f_y(x,y) = \frac{\partial f}{\partial y} = b$ $\quad \nabla f(x,y) = (a, b)$

Compute the gradient vector for a hyperplane (0.5 point)

- Note: derivative of the sum is the sum of the derivatives

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

$f_{x_1}(x_1, x_2, \cdots, x_N) = \frac{\partial f}{\partial x_1} = a_1$, $f_{x_2}(x_1, x_2, \cdots x_N) = \frac{\partial f}{\partial x_2} = a_2$

$\cdots\cdots$, $f_{x_N}(x_1, x_2, \cdots, x_N) = \frac{\partial f}{\partial x_N} = a_N$

$\Rightarrow \nabla f(x) = a_1 + a_2 + \cdots + a_N$

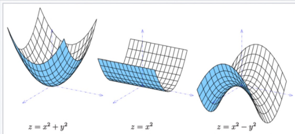
- Compute the partial derivative of the paraboloid function (1.5 point)

$$z = f(x, y) = A(x - x_o)^2 + B(y - y_o)^2 + C$$

$$f_x(x, y) = \left(\frac{\partial f(x, y)}{\partial x}\right)_y = ?$$

$$f_y(x, y) = \left(\frac{\partial f(x, y)}{\partial y}\right)_x = ?$$

Aside: Types of paraboloids



elliptic paraboloid, parabolic cylinder, hyperbolic paraboloid

$$f_x(x,y) = \left(\frac{\partial f(x,y)}{\partial x}\right)_y = \frac{\partial(Ax^2 - 2AX \cdot X_0 + AX_0^2)}{\partial X}$$

$$= 2AX - 2AX_0$$

$$f_y(x,y) = \left(\frac{\partial f(x,y)}{\partial y}\right)_x = \frac{\partial(By^2 - 2By \cdot y_0 + By_0^2)}{\partial y}$$

$$= 2By - 2By_0$$

- Given the following matrices and vectors (1.5 point)

$$\mathbf{x} = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} 2 & 5 & 1 \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 3 & 5 \\ 5 & 2 \\ 1 & 4 \end{pmatrix}$$

$[3 \times 1]$ $\quad [1 \times 3]$ $\quad [3 \times 3]$ $\quad [3 \times 2]$

- compute the following quantities and specify the shape of the output, if an operation is not defined then just say "not defined"
  - (where dot specifies a dot product and x specifies a matrix product).

$\mathbf{x}^T$ $\quad \mathbf{y}^T$ $\quad \mathbf{B}^T$ $\quad \mathbf{x} \cdot \mathbf{x}$ $\quad \mathbf{x} \cdot \mathbf{y}^T$ $\quad \mathbf{x} \times \mathbf{y}$ $\quad \mathbf{y} \times \mathbf{x}$ $\quad \mathbf{A} \times \mathbf{x}$ $\quad \mathbf{A} \times \mathbf{B}$ $\quad$ **B.reshape(1,6)**

$$X^T = (3 \ 1 \ 4) \qquad y^T = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}$$

$[1 \times 3]$ $\qquad [3 \times 1]$

$$X \cdot X = X^T \times X = \underset{[1\times3]}{(3 \ 1 \ 4)} \times \underset{[3\times1]}{\begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}} = \underset{[1\times1]}{3\times3+1\times1+4\times4 = 26}$$

$$X \cdot y^T = X^T \times y^T = \underset{[1\times3]}{(3 \ 1 \ 4)} \times \underset{[3\times1]}{\begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}} = \underset{[1\times1]}{3\times2+1\times5+4\times1 = 15}$$

$$A \times X = \underset{[3\times3]}{\begin{pmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{pmatrix}} \times \underset{[3\times1]}{\begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}} \Rightarrow [3\times1]$$

$$= \begin{pmatrix} 4\times3+5\times1+2\times4 \\ 3\times3+1\times1+5\times4 \\ 6\times3+4\times1+3\times4 \end{pmatrix} = \underset{[3\times1]}{\begin{pmatrix} 25 \\ 30 \\ 34 \end{pmatrix}}$$

$$A \times B = \underset{[3\times3]}{\begin{pmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{pmatrix}} \times \underset{[3\times2]}{\begin{pmatrix} 3 & 5 \\ 5 & 2 \\ 1 & 4 \end{pmatrix}} \Rightarrow [3\times2]$$

$$= \begin{pmatrix} 4\times3+5\times5+2\times1 & 4\times5+5\times2+2\times4 \\ 3\times3+1\times5+5\times1 & 3\times5+1\times2+5\times4 \\ 6\times3+4\times5+3\times1 & 6\times5+4\times2+3\times4 \end{pmatrix}$$

$$= \begin{pmatrix} 39 & 38 \\ 19 & 37 \\ 41 & 50 \end{pmatrix}$$

B.reshape (1,6)

$$B = \begin{pmatrix} 3 & 5 \\ 5 & 2 \\ 1 & 4 \end{pmatrix} \Rightarrow \begin{bmatrix} 3 & 5 & 5 & 2 & 1 & 4 \end{bmatrix}$$

$[3\times2] \Rightarrow [1\times6]$

Linear least squares (LLS): Single-variable (6 points)

- 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)

Model: $y = M(x\,|\,\mathbf{p}) = mx + b$

$\mathbf{p} = (p_0, p_1) = (m, b)$

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

solution: $m = \dfrac{cov(x,y)}{var(x)}$ $\quad \bar{x} = \dfrac{1}{N}\sum_{i=1}^{N} x_i \quad \mathrm{var}(X) = \dfrac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$b = \bar{y} - \dfrac{cov(x,y)}{var(x)}\bar{x} \quad \bar{y} = \dfrac{1}{N}\sum_{i=1}^{N} y_i \quad \mathrm{cov}(X,Y) = \dfrac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$y = M(x|p) = mx + b \quad p = (p_0, p_1) = (m, b)$

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

Suppose $L = \sum_{i=1}^{n} (y_i - mx_i - b)^2$ In order to minimize L:

$\Rightarrow \dfrac{\partial L}{\partial m} = 0 \Rightarrow -2\left(\sum_{i=1}^{n}(y_i - mx_i - b)\right) = 0 \Rightarrow \sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i - nb = 0$ ①

$\Rightarrow \dfrac{\partial L}{\partial b} = 0 \Rightarrow -2\left[\sum_{i=1}^{n}(y_i - mx_i - b)x_i\right] = 0 \Rightarrow \sum_{i=1}^{n} x_i y_i - m\sum_{i=1}^{n} x_i^2 - b\sum_{i=1}^{n} x_i = 0$ ②

From ① $\Rightarrow b = \dfrac{\sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i}{n} = \dfrac{\sum_{i=1}^{n} y_i}{n} - m\dfrac{\sum_{i=1}^{n} x_i}{n}$ ③ $= \bar{y} - m\bar{x}$ ④

plug ③ into ② $\Rightarrow \sum_{i=1}^{n} x_i y_i - m\sum_{i=1}^{n} x_i^2 - \left(\dfrac{\sum_{i=1}^{n} y_i}{n} - m\dfrac{\sum_{i=1}^{n} x_i}{n}\right)\sum_{i=1}^{n} x_i = 0$

$\Rightarrow \sum_{i=1}^{n} x_i y_i - m\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + \frac{1}{n}\cdot m\cdot\left(\sum_{i=1}^{n} x_i\right)^2 = 0$

$\Rightarrow \sum_{i=1}^{n} x_i\left(\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\cdot m\sum_{i=1}^{n} x_i\right) + m\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$

$\Rightarrow m\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]^2\right] = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$

$\Rightarrow m = \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]^2}$ ⑤ ⑥

For ⑤ $\Rightarrow \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \times n\bar{x} \times n\bar{y} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$

$= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}$

$= \sum_{i=1}^{n}\left(x_i y_i - x_i\bar{y} - y_i\bar{x} + \bar{x}\bar{y}\right)$

$= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

For ⑥ $\Rightarrow \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\cdot n\bar{x}\cdot n\bar{x} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$

$\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 = \sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum_{i=1}^{n}(x_i - \bar{x})^2$

Therefore, $m = \dfrac{⑤}{⑥} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{\mathrm{cov}(x,y)}{\mathrm{Var}(x)}$

From ④ $\Rightarrow b = \bar{y} - m\bar{x} = \bar{y} - \dfrac{\mathrm{cov}(x,y)}{\mathrm{Var}(x)}\bar{x}$

Linear least squares (LLS): Multi-variable (EXTRA CREDIT) (+3 points)

- Use matrix calculus to analytically derive the expression for **two variable** linear regression fitting parameters using the sum of square error as the loss function
  - Show your work using matrix notation
  - From your solution infer the generalized solution for an arbitrary number of variables

solution: $\vec{w} = (X^\top X)^{-1} X^\top Y.$

For two variable linear regression :

$$Y = X\beta + E$$

$$\Rightarrow \underset{[n\times 1]}{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}} = \underset{[n\times 3]}{\begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}} \times \underset{[3\times 1]}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}} + \underset{[n\times 1]}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}$$

$E = Y - \hat{Y}$, using the sum of square error :

$$Q = E'E = \underset{[1\times n]}{(\varepsilon_1 \ \varepsilon_2 \cdots \varepsilon_n)} \underset{[n\times 1]}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}}$$

$$= \sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (m + b_1 x_{i1} + b_2 x_{i2})]^2$$

$$\Rightarrow Q = E'E = (Y-\hat{Y})^T (Y-\hat{Y}) = (Y - WX)^T (Y - WX)$$

$$= Y^T Y - WXY^T - W^T X^T Y + W^T X^T \cdot X \cdot W$$

$$= Y^T Y - 2Y^T X W + W^T W X^T X$$

minimize $\Rightarrow \frac{\partial Q}{\partial W} = -2Y^T X + 2WX^T X = 0$

$$= -2(Y^T X)^T + WX^T X = 0$$

$$\Rightarrow X^T X W = X^T Y$$

$$\Rightarrow \vec{W} = (X^T X)^{-1} X^T Y$$