Part-1: Analytic assignment (10 points)

i.e. with pen and paper (show your work and all steps)

Note:

$\frac{d}{dx}x^2 = 2x$

$\frac{d}{dx}x^n = nx^{n-1}$

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

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

- Compute the gradient vector for a hyperplane (0.5 point)
  - Note: derivative of the sum is the sum of the derivatives

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

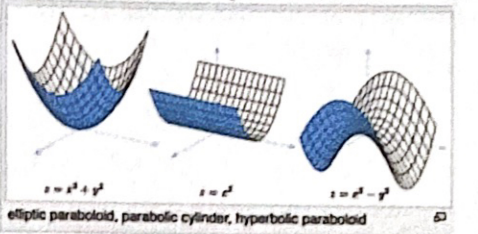
- Compute the partial derivative of the paraboloid function (1.5 point)

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

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

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

Aside: Types of paraboloids



elliptic paraboloid, parabolic cylinder, hyperbolic paraboloid

- 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).

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

- $\nabla f(x,y) = (f_x(x,y), f_y(x,y)) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \Rightarrow \frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(ax+by+c) = a$ $\Rightarrow (a, b)$

  $\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(ax+by+c) = b$

- $\nabla f(x) = (f_x(x)) = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left(\sum_{i=1}^{N} a_i(x_i - b_i) + S\right) = (a_1 + a_2 + \cdots + a_n) = \left(\sum_{i=1}^{N} a_i\right)$

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

  $f_x(x,y) = \left(\frac{\partial f(x,y)}{\partial x}\right)_y = \frac{\partial}{\partial x}\left(A(x-x_0)^2 + B(y-y_0)^2 + C\right) = 2A(x-x_0)$

  $f_y(x,y) = \left(\frac{\partial f(x,y)}{\partial y}\right)_x = \frac{\partial}{\partial y}\left(A(x-x_0)^2 + B(y-y_0)^2 + C\right) = 2B(y-y_0)$

- ① $x^T = [3 \ 1 \ 4]$, shape $[1\times3]$ ② $y^T = \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$, shape $[3\times1]$

③ $B^T = \begin{bmatrix} 3 & 5 & 1 \\ 5 & 2 & 4 \end{bmatrix}$, shape $[2\times3]$ ④ $x \cdot x = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}$, undefined $[3\times1]\cdot[3\times1]$

⑤ $x \cdot y^T = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix}$, undefined $[3\times1]\cdot[3\times1]$

⑥ $x \times y = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \times [2 \ 5 \ 1] = \begin{bmatrix} 3(2) & 3(5) & 3(1) \\ 1(2) & 1(5) & 1(1) \\ 4(2) & 4(5) & 4(1) \end{bmatrix} = \begin{bmatrix} 6 & 15 & 3 \\ 2 & 5 & 1 \\ 8 & 20 & 4 \end{bmatrix}$, shape $[3\times3]$

⑦ $y \times x = [2 \ 5 \ 1] \times \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} = [2(3) \ \ 5(1) \ \ 1(4)] = [6 \ 5 \ 4]$, shape $[1\times1]$

⑧ $A \times x = \begin{bmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{bmatrix} \times \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 4(3) + 5(1) + 2(4) \\ 3(3) + 1(1) + 5(4) \\ 6(3) + 4(1) + 3(4) \end{bmatrix} = \begin{bmatrix} 25 \\ 30 \\ 34 \end{bmatrix}$, shape $[3\times1]$

⑨ $A \times B = \begin{bmatrix} 4 & 5 & 2 \\ 3 & 1 & 5 \\ 6 & 4 & 3 \end{bmatrix} \times \begin{bmatrix} 3 & 5 \\ 5 & 2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4(3) + 5(5) + 2(1) & 4(5) + 5(2) + 2(4) \\ 3(3) + 1(5) + 5(1) & 3(5) + 1(2) + 5(4) \\ 6(3) + 4(5) + 3(1) & 6(5) + 4(2) + 3(4) \end{bmatrix} = \begin{bmatrix} 39 & 38 \\ 19 & 37 \\ 41 & 50 \end{bmatrix}$, shape $[3\times2]$

⑩ B.reshape(1,6) $= [3 \ 5 \ 5 \ 2 \ 1 \ 4]$, shape $[1\times6]$

Part-1: Analytic assignment

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

solution:

$$m = \frac{cov(x,y)}{var(x)} \qquad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \text{var}(X) = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

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

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

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

To find b:

$$\frac{\partial L}{\partial b}\left[\sum_{i=1}^{N} (\hat{y}_i - b - mx_i)^2\right]$$

$$0/2 = \sum_{i=1}^{N} -2(\hat{y}_i - b - mx_i)/2$$

$$0 = \sum_{i=1}^{N} (\hat{y}_i - b - mx_i) = \sum_{i=1}^{N} \hat{y}_i - \sum_{i=1}^{N} b - m\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} \hat{y}_i - Nb - m\sum_{i=1}^{N} x_i$$

$$b = \frac{\sum_{i=1}^{N} \hat{y}_i - m\sum_{i=1}^{N} x_i}{N} = \bar{y} - m\bar{x} = \boxed{\bar{y} - \frac{cov(x,y)}{var(x)}\bar{x} = b}$$

To find m:

$$\frac{\partial L}{\partial m}\left[\sum_{i=1}^{N} (\hat{y}_i - b - mx_i)^2\right]$$

$$0/2 = \sum_{i=1}^{N} -2x_i(\hat{y}_i - b - mx_i)/2$$

$$0 = \sum_{i=1}^{N} x_i(\hat{y}_i - b - mx_i) = \sum_{i=1}^{N} (y_i x_i - bx_i - mx_i^2)$$

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

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

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

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

$$=$$

$$\frac{cov(x,y)}{var(x)} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \sum_{i=1}^{N} \frac{y_i - \bar{y}}{x_i - \bar{x}} \quad \therefore \quad \boxed{m = \frac{cov(x,y)}{var(x)}}$$

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

Model: $y_i = \vec{w}^T \vec{x}_i$

Loss surface: (Squared Error) $SE = \sum_{i=1}^{N} (y_i - \vec{x}_i \cdot \vec{w})^2$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} - & \vec{x}_1 & - \\ & \vdots & \\ - & \vec{x}_n & - \end{bmatrix}$$

$$SE = \| Y - X\vec{w} \|^2$$

$$\frac{\partial}{\partial \vec{w}} SE = \frac{\partial}{\partial \vec{w}} (Y - X\vec{w})^T (Y - X\vec{w})$$

$$= \frac{\partial}{\partial \vec{w}} (Y^T Y - 2\vec{w}^T X^T + \vec{w}^T X^T X Y)$$

$$0 = -2X^T Y + 2X^T X \vec{w}$$

$$X^T X \vec{w} = X^T Y \longrightarrow \vec{w} = \frac{X^T Y}{X^T X}$$

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