HW 2.1: Analytic Assignment

Sam Solomon

September 21, 2021

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

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

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x}(ax+by+c) \\ \frac{\partial}{\partial y}(ax+by+c) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x}ax + \frac{\partial}{\partial x}by + \frac{\partial}{\partial y}c \\ \frac{\partial}{\partial y}ax + \frac{\partial}{\partial y}by + \frac{\partial}{\partial y}c \end{bmatrix} = \begin{bmatrix} a+0+0 \\ 0+b+0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}$$

- Compute the gradient vector for a hyperplane (0.5 points)

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

$$\nabla f(\vec{x}) = \nabla f(x_1, x_2, ..., x_N) = \begin{bmatrix} \frac{\partial}{\partial x_1}(a_1x_1 + a_2x_2 + ... + a_Nx_N + d) \\ \frac{\partial}{\partial x_2}(a_1x_1 + a_2x_2 + ... + a_Nx_N + d) \\ \vdots \\ \frac{\partial}{\partial x_N}(a_1x_1 + a_2x_2 + ... + a_Nx_N + d) \end{bmatrix} =$$

$$\begin{bmatrix} \frac{\partial}{\partial x_1}a_1x_1 + \frac{\partial}{\partial x_1}a_2x_2 + ... + \frac{\partial}{\partial x_1}a_Nx_N + \frac{\partial}{\partial x_1}d \\ \frac{\partial}{\partial x_2}a_1x_1 + \frac{\partial}{\partial x_2}a_2x_2 + ... + \frac{\partial}{\partial x_2}a_Nx_N + \frac{\partial}{\partial x_2}d \\ \vdots \\ \frac{\partial}{\partial x_N}a_1x_1 + \frac{\partial}{\partial x_N}a_2x_2 + ... + \frac{\partial}{\partial x_N}a_Nx_N + \frac{\partial}{\partial x_N}d \end{bmatrix} = \begin{bmatrix} a_1 + 0 + ... + 0 + 0 \\ 0 + a_2 + ... + 0 + 0 \\ \vdots \\ 0 + 0 + ... + a_N + 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \vdots \\ \boldsymbol{a_N} \end{bmatrix}$$

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

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

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

$$\frac{\partial}{\partial x}\big(A(x^2 - 2x_0x + x_0^2) + B(y^2 - 2y_0y + y_0^2) + C\big) =$$

$$\frac{\partial}{\partial x}(Ax^2 - 2Ax_0x + Ax_0^2 + By^2 - 2By_0y + By_0^2 + C) =$$

$$\frac{\partial}{\partial x}(Ax^2) - \frac{\partial}{\partial x}(2Ax_0x) + \frac{\partial}{\partial x}(Ax_0^2) + \frac{\partial}{\partial x}(By^2) - \frac{\partial}{\partial x}(2By_0y) + \frac{\partial}{\partial x}(By_0^2) + \frac{\partial}{\partial x}(C) =$$
$$2Ax - 2Ax_0 + 0 + 0 - 0 + 0 + 0 = 2Ax - 2Ax_0$$

$$f_x(x,y) = \left(\frac{\partial f(x,y)}{\partial x}\right) = \boldsymbol{2Ax - 2Ax_0}$$

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

$$\frac{\partial}{\partial x}\big(A(x^2 - 2x_0x + x_0^2) + B(y^2 - 2y_0y + y_0^2) + C\big) =$$

$$\frac{\partial}{\partial y}(Ax^2 - 2Ax_0x + Ax_0^2 + By^2 - 2By_0y + By_0^2 + C) =$$

$$\frac{\partial}{\partial y}(Ax^2) - \frac{\partial}{\partial y}(2Ax_0x) + \frac{\partial}{\partial y}(Ax_0^2) + \frac{\partial}{\partial y}(By^2) - \frac{\partial}{\partial y}(2By_0y) + \frac{\partial}{\partial y}(By_0^2) + \frac{\partial}{\partial y}(C) =$$
$$0 - 0 + 0 + 2By - 2By_0 + 0 + 0 = 2By - 2By_0$$

$$f_y(x,y) = \left(\frac{\partial f(x,y)}{\partial y}\right) = \boldsymbol{2By - 2By_0}$$

- Given the following matrices and vectors (1.5 points)

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

compute the following quantities and specify the shape of the output. If an operation is not defined then just say "not defined".

$$\boldsymbol{x^T} = \begin{pmatrix} \mathbf{3} & \mathbf{1} & \mathbf{4} \end{pmatrix}$$

$$\boldsymbol{y^T} = \begin{pmatrix} \mathbf{2} \\ \mathbf{5} \\ \mathbf{1} \end{pmatrix}$$

$$\boldsymbol{B^T} = \begin{pmatrix} \mathbf{3} & \mathbf{5} & \mathbf{1} \\ \mathbf{5} & \mathbf{2} & \mathbf{4} \end{pmatrix}$$

$$\boldsymbol{x} \cdot \boldsymbol{x} \textbf{ is not defined}$$

$$\boldsymbol{x} \cdot \boldsymbol{y^T} \textbf{ is not defined}$$

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

$$\boldsymbol{y} \times \boldsymbol{x} = \begin{pmatrix} 3(2) + 1(5) + 4(1) \end{pmatrix} = \begin{pmatrix} 6 + 5 + 4 \end{pmatrix} = \begin{pmatrix} \mathbf{15} \end{pmatrix}$$

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

$$\boldsymbol{A} \times \boldsymbol{B} = \begin{pmatrix} 4(3) + 5(5) + 2(1) & 4(5) + 5(2) + 2(4) \\ 3(3) + 1(5) + 5(1) & 3(5) + 1(2) + 5(4) \\ 6(3) + 4(5) + 3(1) & 6(5) + 4(2) + 3(4) \end{pmatrix} =$$

$$\begin{pmatrix} 12 + 25 + 2 & 20 + 10 + 8 \\ 9 + 5 + 5 & 15 + 2 + 20 \\ 18 + 20 + 3 & 30 + 8 + 12 \end{pmatrix} = \begin{pmatrix} \mathbf{39} & \mathbf{38} \\ \mathbf{19} & \mathbf{37} \\ \mathbf{41} & \mathbf{50} \end{pmatrix}$$

$$\boldsymbol{B}.\text{reshape(1,6)} = \begin{pmatrix} \mathbf{3} & \mathbf{5} & \mathbf{5} & \mathbf{2} & \mathbf{1} & \mathbf{4} \end{pmatrix}$$

- Use calculus to analytically derive the expression for single variable linear regression fitting parameters using the sum of square error as the loss function.

  Model: $y = M(x|\vec{p}) = mx + b$, such that $\vec{p} = (p_0, p_1) = (m, b)$
  Loss surface: $L(\vec{p}) = L(m, b) = \sum_{i=1}^{n}((\hat{y}_i - M(\hat{x}_i, m, b))^2 =$

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

  The sum of square error is minimized when the gradient of the loss function is 0, i.e., when $\nabla L(\vec{p}) = \left(\frac{\partial L}{\partial m}, \frac{\partial L}{\partial b}\right) = \vec{0}$, or when

  $$\frac{\partial L}{\partial m} = \frac{\partial}{\partial m}\sum_{i=1}^{n}(\hat{y}_i - m\hat{x}_i - b)^2 = \sum_{i=1}^{n}\frac{\partial}{\partial m}(\hat{y}_i - m\hat{x}_i - b)^2 = 0$$

  and

  $$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b}\sum_{i=1}^{n}(\hat{y}_i - m\hat{x}_i - b)^2 = \sum_{i=1}^{n}\frac{\partial}{\partial b}(\hat{y}_i - m\hat{x}_i - b)^2 = 0$$

  When $\frac{\partial L}{\partial m} = 0$,

  $$\sum_{i=1}^{n}2(\hat{y}_i - m\hat{x}_i - b)(-\hat{x}_i) = 0 \text{ (by the chain rule)}$$

  so

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

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

When $\dfrac{\partial L}{\partial b} = 0$,

$$\sum_{i=1}^{n} 2(\hat{y}_i - m\hat{x}_i - b)(-1) = 0 \text{ (by the chain rule)}$$

so

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

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

$$nb = \sum_{i=1}^{n}\hat{y}_i - \sum_{i=1}^{n}m\hat{x}_i$$

$$b = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}m\hat{x}_i = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i - (m)\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i = \bar{y} - m\bar{x}$$

$$\text{where } \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i$$

Recall that when $\dfrac{\partial L}{\partial m} = 0$, $\sum_{i=1}^{n}\hat{x}_i\hat{y}_i - m\sum_{i=1}^{n}\hat{x}_i^2 - b\sum_{i=1}^{n}\hat{x}_i = 0$.
Therefore, when $\dfrac{\partial L}{\partial m} = 0$ and $\dfrac{\partial L}{\partial b} = 0$,

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

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

$$\sum_{i=1}^{n}\hat{x}_i\hat{y}_i - m\sum_{i=1}^{n}\hat{x}_i^2 - \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i\sum_{i=1}^{n}y_i + m\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2 = 0$$

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

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

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

$$COV(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i\hat{y}_i - \hat{x}_i\bar{y} - \hat{y}_i\bar{x} + \bar{x}\bar{y}) =$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i\hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i\bar{y} - \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i\bar{x} + \frac{1}{n}\sum_{i=1}^{n}\bar{x}\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i\hat{y}_i - \frac{1}{n}(n\bar{x})\bar{y} - \frac{1}{n}(n\bar{y})\bar{x} + \frac{1}{n}(n\bar{x}\bar{y}) =$$

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

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

$$VAR(x) = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i^2 - 2\hat{x}_i\bar{x} + \bar{x}^2) = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i + \frac{1}{n}\sum_{i=1}^{n}\bar{x}^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - 2\bar{x}\frac{1}{n}(n\bar{x}) + \frac{1}{n}(n\bar{x}^2) = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i\right)^2 =$$

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

It follows that when $\dfrac{\partial L}{\partial m} = 0$ and $\dfrac{\partial L}{\partial b} = 0$

$$m = \frac{\sum_{i=1}^{n} \hat{x}_i \hat{y}_i - \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i \sum_{i=1}^{n} \hat{y}_i}{\sum_{i=1}^{n} \hat{x}_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} \hat{x}_i\right)^2} = \frac{\left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} \hat{x}_i \hat{y}_i - \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i \sum_{i=1}^{n} \hat{y}_i\right)}{\left(\frac{1}{n}\right)\left(\sum_{i=1}^{n} \hat{x}_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} \hat{x}_i\right)^2\right)} = \frac{COV(x,y)}{VAR(x)}$$

and therefore

$$\boldsymbol{m = \frac{COV(x,y)}{VAR(x)}}$$

Recall that $b = \bar{y} - m\bar{x}$ when $\dfrac{\partial L}{\partial b} = 0$. Therefore, when $\dfrac{\partial L}{\partial m} = \dfrac{\partial L}{\partial b} = 0$,

$$\boldsymbol{b = \bar{y} - \frac{COV(x,y)}{VAR(x)}\bar{x}}$$

For approach to problem-solving, the following source was consulted:

"Least squares: Calculus to find residual minimizers?", StackExchange, accessed September 18, 2021, https://stats.stackexchange.com/questions/133554/least-squares-calculus-to-find-residual-minimizers