AI535 HW1

JuHyun Kim

1. (Optimization) Compute the gradient $\nabla f(\mathbf{x})$ and Hessian $\nabla^2 f(\mathbf{x})$ of the function (5 points)

$f(\mathbf{x}) = (x_1 + x_2)(x_1 x_2 + x_1 x_2^2)$

Find at least 3 stationary points of this function (3 points). Show that $[3/8, -6/8]^\top$ a local maximum of this function (2 point).

$$f(x) = (x_1+x_2)(x_1x_2+x_1x_2^2)$$
$$= x_1^2x_2 + x_1^2x_2^2 + x_1x_2^2 + x_1x_2^3$$

① $\nabla f(x) = \begin{bmatrix} \frac{df}{dx_1} = 2x_1x_2 + 2x_1x_2^2 + x_2^2 + x_2^3 \\ \frac{df}{dx_2} = x_1^2 + 2x_1^2x_2 + 2x_1x_2 + 3x_1x_2^2 \end{bmatrix}$

Hessian $\nabla^2 f(x) = \begin{bmatrix} \frac{d^2f}{dx_1^2} & \frac{d^2f}{dx_1x_2} \\ \frac{d^2f}{dx_1x_2} & \frac{d^2f}{dx_2^2} \end{bmatrix} = \begin{bmatrix} 2x_2+2x_2^2 & 2x_1+4x_1x_2+2x_2+3x_2^2 \\ 2x_1+4x_1x_2+2x_2+3x_2^2 & 2x_1^2+2x_1+6x_1x_2 \end{bmatrix}$

② Stationary Points → $(0,0)$ $(1,-1)$ $(0,-1)$

③ Show $[\frac{3}{8}, -\frac{6}{8}]^T$ a local maximum

$$\nabla^2 f(x_1 = \tfrac{3}{8}, x_2 = -\tfrac{6}{8}) = \begin{bmatrix} 2(-\frac{6}{8}) + 2(-\frac{6}{8})^2 & 2(\frac{3}{8}) + 4(\frac{3}{8})\cdot(-\frac{6}{8}) + 2(-\frac{6}{8}) + 3(-\frac{6}{8})^2 \\ // & 2(\frac{3}{8})^2 + 2\cdot\frac{3}{8} + 6\cdot\frac{3}{8}(-\frac{6}{8}) \end{bmatrix}$$

$$= \begin{bmatrix} -0.375 & -0.1875 \\ -0.1875 & -0.65625 \end{bmatrix}$$

$$\det(H) = (-0.375)(-0.65625) - (-0.1875)^2$$

$$= 0.246 - 0.035 = 0.211 > 0 \longrightarrow \text{it is minima or maxima}$$

Since $\nabla^2 f(\frac{3}{8})$ and $\nabla^2 f(-\frac{6}{8})$ are smaller then zero,

it is local maximum of this function.

2. (Optimization) Show that the function $f(\mathbf{x}) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$ has only one stationary point (4 points), and that it is neither a minimum nor a maximum, but is a saddle point (4 points).

① Show $f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$ has only one stationary point.

⇒ To get stationary point, have to find the x_1, x_2 value where $f'(x) = 0$.

$f'(x_1) = 8 + 2x_1 = 0 \qquad x_1 = -4$

$f'(x_2) = 12 - 4x_2 = 0 \qquad x_2 = 3$

$[-4 \;\; 3]^T$ is the only stationary point of $f(x)$.

② Show $[-4 \;\; 3]^T$ is a saddle point.

⇒ If $\det(H(x)) < 0$, x is saddle point.

$f''(x_1) = 2$

$f''(x_2) = -4$

$H = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix} \qquad \det(H) = 2(-4) - 0) = -8$

Since $\det(H) = -8$ is smaller then zero, $[-4 \;\; 3]^T$ is the saddle point.

3. (Linear Algebra) If $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices, prove that the matrix $\begin{bmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B} \end{bmatrix}$ is also positive definite (7 points).

⇒ Using Definition of Positive Definiteness

For symmetric matrix M, we have $z^T M z > 0$ for every z.

Let $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = M$, $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ (where z_1 is a vector of same size as A and z_2 is a vector of same size as B)

$[z_1 \;\; z_2]^T \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = z_1^T A z_1 + z_2^T B z_2$

Since A and B are positive define, we can says $z_1^T A z_1 > 0$ and $z_2^T B z_2 > 0$.

Therefore, $z_1^T A z_1 + z_2^T B z_2$ is also bigger than zero, which proves that $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is positive definite.

4. (Chain Rule Calculus) Consider this function: $f(\mathbf{x}) = \mathbf{w}_2^\top sigmoid(\mathbf{W}_1\mathbf{x})$, where $sigmoid(x) = \frac{1}{1+e^{-x}}$ applies to each entry of the vector, please compute the derivatives of $\frac{\partial f}{\partial \mathbf{w}_2}, \frac{\partial f}{\partial \mathbf{W}_1}, \frac{\partial f}{\partial \mathbf{x}}$ (15 points), $\mathbf{W}_1$ is $c \times d$, $\mathbf{x}$ is $d \times 1$, $\mathbf{w}_2$ is $c \times 1$.

$$f(x) = W_2^T \cdot \frac{1}{1+e^{-(W_1 x)}}$$

$$sigmoid(W_1 x) = [\quad]_{c \times d} [\quad]_{d \times 1} = c \times 1$$

① $$\frac{df}{dW_2} = \frac{1}{1+e^{-(W_1 x)}} \quad [c \times 1]$$

② $$\frac{df}{dW_1} = W_2^T \cdot \frac{1}{(1+e^{-(W_1 x)})^2} (-e^{-(W_1 x)})(-x^T)$$

$$(1 \times c) \quad (c \times 1) \quad (c \times 1)(1 \times d) \Rightarrow [c \times d]$$

③ $$\frac{df}{dx} = W_2^T \cdot \frac{1}{(1+e^{-(W_1 x)})^2} (-e^{-(W_1 x)})^T (-W_1)$$

$$(1 \times c) \quad (c \times 1) \quad (1 \times c)(c \times d) \Rightarrow [1 \times d]$$

3/4

5. (High Dimensional Statistics ("Curse of Dimensionality")) Consider N data points independent and uniformly distributed in a p-dimensional unit ball B (for every $x \in B, \|x\|^2 \leq 1$), centered at the origin. The median distance from the origin to the closest data point is given by the expression:

$$d(p, N) = \left(1 - \frac{1}{2}^{\frac{1}{N}}\right)^{\frac{1}{p}}$$

Prove this expression (8 points). Compute the median distance $d(p, N)$ for $N = 10,000, p = 1,000$ (2 points).

Hint: The volume of a ball in p dimensions is $V_p(R) = \frac{\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}+1\right)} R^p$, where R is the radius of the ball, and Γ is the Gamma function (the exact form of it does not matter for this assignment). A point being the **closest** point to the origin means that there is **no** point that has a smaller distance to the origin than itself. What is the **probability** for that to happen with a uniform distribution in a unit ball?

Median distance (d) is the distance at which the probability that any point is closer to the origin than this distance is $\frac{1}{2}$.

For ball B, R = 1, the volume is $V_p(1) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2}+1)} \cdot 1$

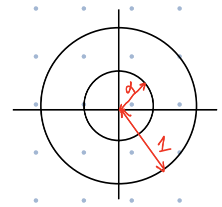
the volume of ball which R = d is $V_p(d) = \frac{\pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2}+1)} \cdot d^p$

And each data points are independent, therefore we can prove like below

$$\frac{1}{2} = \prod_{i=1}^{N} [1 - P(\|x_i\| \leq d)]$$

$$\frac{1}{2} = \prod_{i=1}^{N} [1 - d^p]$$

$$\frac{V_p(d)}{V_p(1)} = \frac{\cancel{\textcircled{1}}\, d^p}{\cancel{\textcircled{1}}\, 1} = d^p$$



$$\frac{1}{2} = (1 - d^p)^N$$

$$1 - d^p = \left(\frac{1}{2}\right)^{\frac{1}{N}} \rightarrow d^p = 1 - \left(\frac{1}{2}\right)^{\frac{1}{N}} \rightarrow d = \left[1 - \left(\frac{1}{2}\right)^{\frac{1}{N}}\right]^{\frac{1}{p}}$$

$d(1000, 10000) = 0.9905$

4/4