Sumeet Shanbhag
ID:- 641020714

Deep Learning Proofs.

Q 2. a) $f(x,y) = x^4 + xy + x^2$

Prove that f is a ~~g~~ convex by proving its Hessian matrix is PSD or identify a point in the domain of f where Hessian is not PSD.

The Hessian matrix can be found out by :-

$$H = \begin{bmatrix} \partial^2 f / \partial x \partial x & \partial^2 f / \partial x \partial y \\ \partial^2 f / \partial y \partial x & \partial^2 f / \partial y \partial y \end{bmatrix}$$

$$\frac{\partial f}{\partial x} = 4x^3 + y + 2x \quad , \quad \frac{\partial f}{\partial x \partial x} = 12x^2 + 2$$

$$\frac{\partial f}{\partial y} = x \quad , \quad \frac{\partial f}{\partial y \partial y} = 1.$$

$$\therefore H = \begin{bmatrix} 12x^2 + 2 & 1 \\ 1 & 0 \end{bmatrix}$$

We know that the Hessian ~~of~~ matrix determinant has to be positive or 0 for it to be positive semi definite.

$$\Delta_H = (12x^2 + 2) \times 0 \cancel{+} - 1 \times 1 = -1$$

Since the determinant of the matrix is -1, this means that there exists a point where the Hessian is negative. Thus f is not a convex function.

2b) Show that, for any real vector v, it is always true that $v^T H v \geq 0$.

$$F_{MSE}(w) = \frac{1}{2n} (X^T w - y)(X^T w - y)$$

$$= \frac{1}{2n} ((X^T w - y))^2$$

$$\frac{\partial F_{MSE}(w)}{\partial w} = \frac{1}{\cancel{2}n} \times \cancel{2} \times (x) \times (x^T w - y)$$

$$\frac{\partial F_{MSE}(w)}{\partial w^2} = \frac{1}{\cancel{2}n} \times (X) \times (X^T) = \text{Hessian matrix.} = H$$

To prove that $v^T H v \geq 0$,

$$v^T H v = (1/n) \times (v^T) \times (X) \times (X^T) \times v$$

$X . X^T$ is a positive semi definite matrix which means that $(v^T) * (X) * (X^T) * (v)$ will also be ≥ 0.

This means that H is also PSD, which means the func^n $F_{MSE}(w)$ is convex wrt weight vector w.

4a) $\sigma(x) = \dfrac{1}{1+e^{-x}}$

Prove that $\sigma(-x) = 1 - \sigma(x) \quad \forall x$.

$LHS = \sigma(-x) = \dfrac{1}{1+e^{x}}$

$RHS = 1 - \sigma(x) = 1 - \dfrac{1}{1+e^{-x}}$

$= \dfrac{1+e^{-x} - 1}{1+e^{-x}}$

$= \dfrac{e^{-x}}{1+e^{-x}}$

Multiplying by $\dfrac{e^{x}}{e^{x}}$, we get

$= \dfrac{e^{-x} \times e^{x}}{(1+e^{-x})e^{x}}$

$= \dfrac{1}{1+e^{x}}$

$\therefore LHS = RHS$

4b) Prove that $\sigma'(x) = \frac{\partial \sigma(x)}{\partial x} = \sigma(x)(1-\sigma(x)) \quad \forall x$

$$\frac{\partial \sigma(x)}{\partial x} = \frac{1}{(1+e^{-x})^2} \times e^{-x} \times (-1)$$

$$= \frac{-e^{-x}}{(1+e^{-x})^2} = \text{LHS}.$$

$$\text{RHS} = \sigma(x)(1-\sigma(x))$$

$$= \frac{1}{(1+e^{-x})}\left(1 - \frac{1}{(1+e^{-x})}\right)$$

$$= \frac{1}{(1+e^{-x})}\left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$\therefore$ LHS = RHS