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

$\frac{\partial f}{\partial x} = 4x^3 + y + 2x$; $\frac{\partial f}{\partial y} = x$

$\frac{\partial^2 f}{\partial x^2} = 12x^2 + 2$; $\frac{\partial^2 f}{\partial y^2} = 0$; $\frac{\partial^2 f}{\partial x \partial y} = 1 = \frac{\partial^2 f}{\partial y \partial x}$

$\therefore \quad H = \begin{bmatrix} 12x^2+2 & 1 \\ 1 & 0 \end{bmatrix} \qquad H - \lambda I = \begin{bmatrix} 12x^2+2-\lambda & 1 \\ 1 & -\lambda \end{bmatrix}$

$|H - \lambda I| = -\lambda(12x^2+2-\lambda) - 1 = 0 \Rightarrow \lambda^2 - (12x^2+2)\lambda - 1 = 0$

$\lambda = \frac{12x^2+2 \pm \sqrt{(12x^2+2)^2 - 4 \times (-1)}}{2} = \frac{12x^2+2 \pm \sqrt{(12x^2+2)^2+4}}{2}$

Clearly $\sqrt{(12x^2+2)^2+4} > \sqrt{(12x^2+2)^2} = 12x^2+2 > 0$

$\therefore$ λ can be < 0 for all $x \in \mathbb{R}$

$\therefore$ the matrix is not PSD

CONTINUED ...

Consider the point $(1, -8)$

At this pt, the Hessian is $\begin{bmatrix} 12 \times 1^2 + 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix}$

This matrix is not PSD if $\exists\, v \in \mathbb{R}^2 \mid v^T H v < 0$

Consider $v = \begin{bmatrix} 1 \\ -8 \end{bmatrix}$

$$v^T H v = \begin{bmatrix} 1 & -8 \end{bmatrix} \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -8 \end{bmatrix} = \begin{bmatrix} 1 & -8 \end{bmatrix} \begin{bmatrix} 6 \\ 1 \end{bmatrix} = 6 - 8 = -2 < 0$$

Hence, the Hessian is not PSD at $(1, -8)$

2) b) To Prove Hessian $\left(\frac{1}{2n}(X^T w - y)^T (X^T w - y)\right)$ is PSD

From HW1 $\nabla_x \left((Ax+b)^T (Ax+b)\right) = A^T (Ax+b)$

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

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

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

$$\nabla_w (\nabla_w \text{MSE}) = \frac{1}{2n} \left(\nabla_w (X X^T w) - \nabla_w (X y)\right)$$

$$= \frac{1}{2n} \nabla_w (X X^T w) - 0$$

$$= \frac{1}{2n} X X^T$$

Now we must show that $\frac{1}{2n} X X^T$ is PSD

We can ignore the $\frac{1}{2n}$ as it is simply a +ve scalar

If $X X^T$ is PSD then

$v^T (X X^T) v \geq 0 \quad \forall v \in \mathbb{R}^n$ if $X X^T$ in an $n \times n$ matrix

$\Rightarrow (v^T X)(X^T v) = (X^T v)^T (X^T v) \geq 0$

$X^T v$ is an $n \times 1$ vector

$\therefore (X^T v)^T$ is $1 \times n$

$\therefore (X^T v)^T (X^T v)$ is the sum of squares of the elements of $X^T v$ which are always ≥ 0 assuming $X^T v \in \mathbb{R}^n$

Hence Proved

508-831-5241 (TEL) 508-831-5824 (FAX)

100 Institute Road, Worcester MA 01609-2280 USA
WPI.EDU/+MATH

(84)

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

(a) $\sigma(-x) = 1 - \sigma(x)$

Soln:

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

$$= \frac{1+e^{x}-e^{x}}{1+e^{x}} \quad (+ - e^{x} \text{ in the numerator})$$

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

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

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

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

Thus, $\sigma(-x) = 1 - \sigma(x)$

Hence proved.

(b) $\sigma'(x) = \frac{\partial \sigma}{\partial x}(x) = \sigma(x)(1-\sigma(x))$.

Sol^n: $\sigma(x) = \frac{1}{1+e^{-x}}$ (let $1+e^{-x} = f(x)$) — ①

$$\frac{\partial \sigma}{\partial x} = \frac{\partial}{\partial x}\left(\frac{1}{f(x)}\right)$$

$$= \frac{-1}{f(x)^2} \cdot \frac{\partial f(x)}{\partial x}$$

$$= \frac{-1}{(1+e^{-x})^2} \cdot (-e^{-x}) \text{ — using ①}$$

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

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

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

$$= \sigma(x)(1-\sigma(x))$$

Since $\sigma(x)$ has x as the only variable, $\sigma'(x) = \frac{\partial \sigma}{\partial x}(x)$.

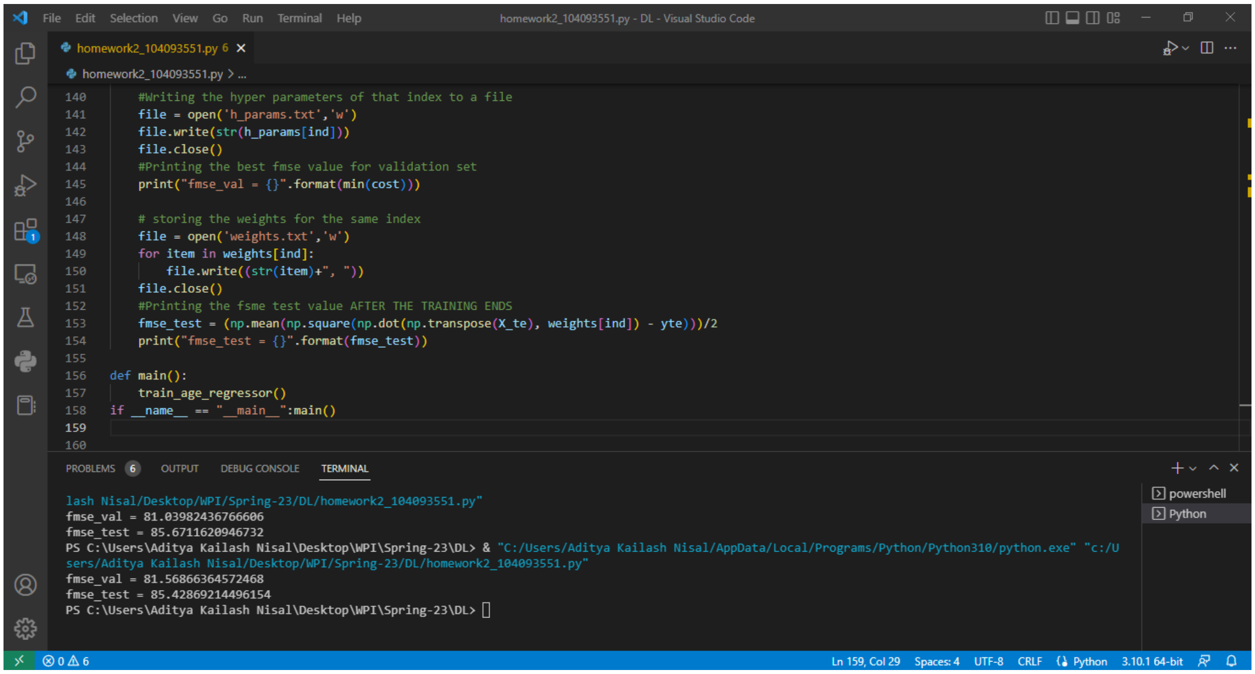
Thus, $\sigma'(x) = \frac{\partial(\sigma(x))}{\partial x} = \sigma(x)(1-\sigma(x))$.

Hence proved.

Homework 2
Deep Learning

Aditya Nisal
Tanish Mishra

Question 3



Reported Fmse on test and validation set

Optimal hyperparameters after testing on the validation set: [50, 0.001, 500, 0.01]

Mini_batch_size = 50
Learning_rate = 0.001
Epochs = 500
Alpha = 0.01

Optimal weights list after training - .txt file included in the submission