

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

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

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

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

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

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

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

$\therefore \lambda$ 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 $\begin{bmatrix} 12x+2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 14 & 1 \\ 1 & 0 \end{bmatrix}$

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

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 HW 1 $\nabla_x ((Ax+b)^T (Ax+b)) = A^T (Ax+b)$

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

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

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

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

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

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

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

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

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

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

$$\therefore (x^T v)^T \text{ is } 1 \times n$$

$\therefore (x^T v)^T (x^T v)$ 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:

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

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

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

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

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

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

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

Hence proved.

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

Soln: $\sigma(x) = \frac{1}{1+e^{-x}} \quad \left(\text{let } 1+e^{-x} = f(x) \right) \rightarrow (1)$

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

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

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

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

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

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

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

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

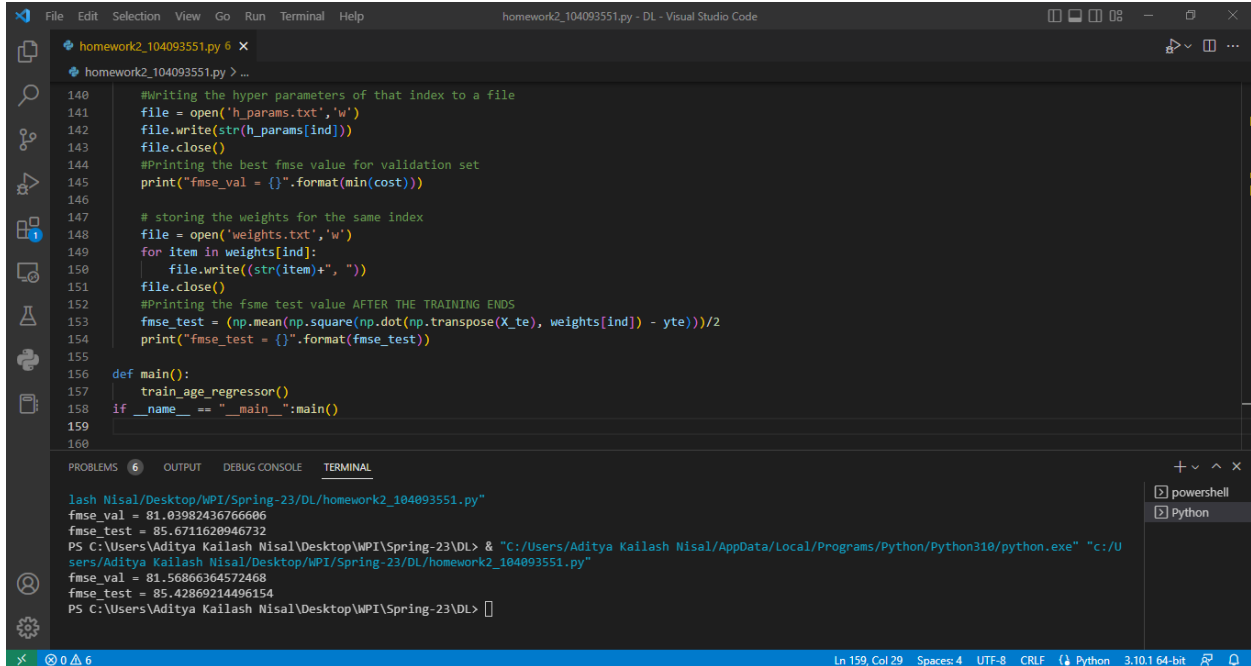
Thus, $\sigma'(x) = \frac{\partial (\sigma(x))}{\partial x} = \sigma(x)(1 - \sigma(x)).$

Hence proved.

Homework 2 Deep Learning

Aditya Nisal
Tanish Mishra

Question 3



The screenshot shows a Visual Studio Code editor with a Python file named `homework2_104093551.py`. The code is a script for training and testing a linear regression model. It includes functions for writing hyperparameters to a file, printing the best validation loss, storing weights, and printing the test loss after training. The `main` function calls `train_age_regressor`. The terminal output shows the execution of the script, displaying the validation loss (`fmse_val`) and test loss (`fmse_test`) for two different sets of hyperparameters.

```
140 #Writing the hyper parameters of that index to a file
141 file = open('h_params.txt','w')
142 file.write(str(h_params[ind]))
143 file.close()
144 #Printing the best fmse value for validation set
145 print("fmse_val = {}".format(min(cost)))
146
147 # storing the weights for the same index
148 file = open('weights.txt','w')
149 for item in weights[ind]:
150     file.write((str(item)+", "))
151 file.close()
152 #Printing the fmse test value AFTER THE TRAINING ENDS
153 fmse_test = (np.mean(np.square(np.dot(np.transpose(X_te), weights[ind]) - yte)))/2
154 print("fmse_test = {}".format(fmse_test))
155
156 def main():
157     train_age_regressor()
158 if __name__ == "__main__":main()
159
160
```

Terminal Output:

```
lash Nisal/Desktop\MPI\Spring-23\DL\homework2_104093551.py"
fmse_val = 81.03982436766606
fmse_test = 85.6711620946732
PS C:\Users\Aditya Kailash Nisal\Desktop\MPI\Spring-23\DL> & "C:/Users/Aditya Kailash Nisal/AppData/Local/Programs/Python/Python310/python.exe" "c:/U
sers/Aditya Kailash Nisal/Desktop\MPI\Spring-23\DL\homework2_104093551.py"
fmse_val = 81.56866364572468
fmse_test = 85.42869214496154
PS C:\Users\Aditya Kailash Nisal\Desktop\MPI\Spring-23\DL>
```

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