CS 1678/2078 Homework 2

Alaa Alghwiri

Written Responses (Part 1)

Given that $f_*(x) = 6x + 4\cos(3x+2) - x^2 + 10\ln(\frac{|x|}{10} + 1) + 7$, find the following:

Problem 1

In this part, I will need to find $\phi(x) = [?]^T$: Based on the given function and in order,

$$\phi(x) = \begin{bmatrix} x & cos(3x+2) & x^2 & \ln(\frac{|x|}{10}+1) & 1 \end{bmatrix}^T$$

Problem 2

In this part i will need to find the optimal weights that corresponds to the features in part 1:

$$w^* = \begin{bmatrix} 6 & 4 & -1 & 10 & 7 \end{bmatrix}^T$$

Problem 3

In this part, i will need to evaluate the same requirements for part 1 and 2 but for the following function:

$$f_*(x) = 6x \times 4\cos(3x+2) \times x^2 \times 10\ln(\frac{|x|}{10} + 1) \times 7$$

In this case, I can simplify the multiplication terms as follow: $6 \times 4 \times 7 \times 10 = 1680$, then we get the following:

$$1680 \times x^3 \times \cos(3x+2) \times \ln(\frac{|x|}{10}+1)$$

$$\phi(x) = \left[x^3 \times \cos(3x+2) \times \ln(\tfrac{|x|}{10}+1)\right]$$
$$w^* = \left[1689\right]^T$$

Problem 4

In this problem, we are looking for:

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

Since differentiation is linear, we can move the derivative inside the summation:

$\frac{1}{2m} \sum_{i=1}^{m} \frac{\partial}{\partial \hat{y}_i} (\hat{y}_i - y_i)^2$

$\frac{\partial}{\partial \hat{y}_i} (\hat{y}_i - y_i)^2 = 2(\hat{y}_i - y_i)$

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

And finally, the derivative evaluates to:

$\frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)$

And this is obviously: $\mathbb{E}[\hat{y} - y]$

In a vector form:

$$\frac{1}{m}\begin{bmatrix}\hat{y_1} - y_1 & \hat{y_2} - y_2 & ... & \hat{y_m} - y_m\end{bmatrix}^T$$

Problem 5

$\hat{y} = f(X, w) = Xw$

Now, i need to find:

$\frac{\partial}{\partial w} f(X, w) = \frac{\partial}{\partial w} \sum_{i=1}^{m} X_i w_i$

Solution:

$= \frac{\partial}{\partial w} \sum_{i=1}^{m} X_i w_i$

$= \sum_{i=1}^{m} \frac{\partial}{\partial w} X_i w_i$

And finally:

$\frac{\partial \hat{y}_i}{\partial w_i} = \sum_{i=1}^{m} X_i$

In a matrix representation:

$$\begin{bmatrix} X_1 & X_2 & X_3 & .... & X_m \end{bmatrix}$$

Problem 6

In this question, i will evaluate the gradient for the loss function xx with respect to the weight w. And this should be expressed in matrices/vectors without summation.

$\nabla l(w) = \frac{\partial}{\partial w} g(f(X, w), y)$

$\frac{\partial}{\partial w} g(f(X, w), y) = \frac{\partial}{\partial w} \frac{1}{2m} \sum_{i=1}^{m} (X_i w_i - y_i)^2$

$= \frac{1}{m} \sum_{i=1}^{m} (X_i w_i - y_i) x_i$

Using matrix representation:

$$\nabla l(w) = X^T (Xw - y)$$

Problem 7

First of all, I will evaluate the gradient of the sigmoid function with respect to the weight w:

$$\frac{\partial f(x,w)}{\partial w} = \frac{\partial}{\partial w}\frac{1}{1+e^{-w^T x}}$$

Using the chain rule and assuming that $u = 1 + e^{-w^T x}$.

$$\frac{\partial f(x,w)}{\partial w} = \frac{\partial}{\partial u}\frac{1}{u} \times \frac{\partial(1+e^{-w^T x})}{\partial w}$$

$$\frac{\partial f(x,w)}{\partial w} = f(x,w)^2 e^{(-w^T x)} X$$

Now, let's evaluate the gradient of the Negative likelihood loss function with respect to the weight w:

$$\nabla l(w) = \frac{\partial}{\partial w} - \sum_{i=1}^{m} y_i \log(f(x_i,w)) + (1-y_i)\log(1-f(x_i,w))$$

$$= -\sum_{i=1}^{m} \frac{y_i}{f(x_i,w)}\frac{\partial f(x_i,w)}{\partial w} - \frac{1-y_i}{1-f(x_i,w)}\frac{\partial f(x_i,w)}{\partial w}$$

$$= -\sum_{i=1}^{m} \frac{y_i}{f(x_i,w)} f(x_i,w)^2 e^{(-w^T x_i)} X_i - \frac{1-y_i}{1-f(x_i,w)} f(x_i,w)^2 e^{(-w^T x_i)} X_i$$

$$= -\sum_{i=1}^{m} y_i(1-f(x_i,w))X_i + (1-y_i)f(x_i,w)X_i$$

$$= -\sum_{i=1}^{m} (y_i - f(x_i,w))X_i$$

$$= X^T(f(X,w) - y)$$