Part -1: Theoretical Questions

Q 1. Provide a complete definition for the following machine learning concepts:

1. Feature vectors
2. Decision boundary
3. Classification
4. Regression
5. Supervised learning
6. Unsupervised learning
7. Overfitting
8. Decision Tree
9. Information Gain
10. Entropy
11. Vanishing gradient

Q2. Suppose we want to predict the price of a house using one feature x , which is the area of the house. Suppose the data points are as follows:

The area x_i (square feet)	The price y_i (dollar)
2,000	250,000
2,500	300,000
3,000	400,000
1,500	220,000
1,000	200,000

Or as follows:

$$\{(x_1, y_1), \quad (x_2, y_2), \quad (x_3, y_3), \quad (x_4, y_4), (x_5, y_5)\}$$
$$= \{(2{,}000, 250{,}000), (2{,}500, 300{,}000), (3{,}000, 400{,}000), (1{,}500, 220{,}000), (1{,}000, 200{,}000)\ \}$$

We use the following linear regression model.

$$\hat{y} = w\, x$$

Consider the square error loss function.

$$J = \frac{1}{2}\,(y - \hat{y})^2 = \frac{1}{2}\sum_{i=1}^{5}(y_i - w\, x_i)^2$$

a) Find the parameter w that optimizes the total loss function J?
   Hint: Optimization problems are solved by taking the derivative of the loss function

with respect to the unknown parameters, setting the derivative to zero, and solving for the unknown parameters.

b) Suppose we are given the area of the house equal to $x = 2{,}100$ Square yard, what is the price of the house $\hat{y}$. Hint: use the value of w that you have estimated in (a).

Q3. Assume we have observations arising from a noisy environment, where the noise is normally distributed.

$$y = \mathbf{w}^T \mathbf{x} + b + \epsilon, \text{ where } \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Thus, we can write the likelihood of seeing a particular y for a given $\mathbf{x}$ via

$$P(y|\mathbf{x}) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} \exp\left(-\frac{1}{2\,\sigma^2}\,(y - \mathbf{w}^T\,\mathbf{x} - b\,)^2\right)$$

Now, according to the principle of maximum likelihood, the best values of parameters $\mathbf{w}$ and b are those that maximize the likelihood of the entire dataset:

$$P(\boldsymbol{y}|\mathbf{X}) = \prod_{i=1}^{n} p(y^{(i)}|\mathbf{x}^{(i)})$$

Maximizing the likelihood is equivalent to minimizing the negative log-likelihood

Write the negative log-likelihood function of this problem $-\log P(\boldsymbol{y}|\mathbf{X})$.

Q4. Summarize the difference between linear regression and logistic regression (using words without equations)

Q5. Suppose we construct a loss function for the logistic regression using the Bernoulli distribution.

We are given N training data with their target labels y_n , where y_n= 0, if $\mathbf{x}_n \in class$ 1 and y=1, if $\mathbf{x}_n \in class$ 2, and n = 1,2, …, N. The probability of success is $\hat{y}$.

$$p(\boldsymbol{y} \mid \mathbf{X}) = \prod_{n=1}^{N} \hat{y}_n^{y_n} (1 - \hat{y}_n)^{(1-y_n)}$$

Write the negative log-likelihood function $-\log p(\boldsymbol{y} \mid \mathbf{X})$.

Q6. What type of problems does the SoftMax function solve? Explain with some details.

Q7. Assume that we have some data $x_1,\ldots,x_n \in$ R. Our goal is to find a constant b such that the loss function $J = \sum_{i=1}^{n}(x_i - b)^2$ is minimized.

Find the analytic solution for the optimal value of b.
Hint: Optimization problems are solved by taking the derivative of the loss function with respect to the unknown parameters, setting the derivative to zero and solving for the unknown parameters.

Part -2: Programming Task:

Build a fully connected neural network (or multi-layer perception) to classify objects in the Fashion MNIST dataset.

To load the data use: fmnist = tf.keras.datasets.fashion_mnist

Please write a summary report (no more than one page) to explain your network architecture and hyper parameters (number of layers, number of neurons in each layer, learning rate choice, number of epochs, etc.), take a screenshot of the output of model.summary() and provide your final training and validation accuracy with a brief summary of your observations.

Please stick to one page.

Please upload your code to your GitHub repository and add only the code link to your report.

Please do NOT copy your entire code to your one-page report.