MACHINE LEARNING CAT 2

1. Explain why you would not use Huber Loss in training Deep FeedForward Networks (2 marks)

Huber Loss is a combination of both Mean Squared Error and Mean Absolute Error. It is hence considered to be less sensitive to outliers. This places Huber loss as a regressor type of function whereas Deep Forward Feed Networks act as Classification functions.

2. One of your AI colleagues is fascinated/confused by a recent research by Facebook AI that uses Convolution Neural Networks for training Natural Language models. Explain why she is fascinated/confused (2 marks)

Convolution Neural Networks is a deep learning algorithm that accepts input through images and assigns learnable weights and biases to various aspects of the image in order to differentiate these aspects from each other.

The fascination of my colleague may be as a result of being impressed by how fast these Convolution Neural Networks are. They have shown some great results whereby they facilitate machines to learn with little to no supervision.

3. According to Ian Goodfellow et al 2016, The core ideas behind modern feedforward networks have not changed substantially since the 1980s. The same back-propagation algorithm and the same approaches to gradient descent are still in use. In a few sentences, discuss/envision what will deep learning look like today if back propagation algorithm was never invented (2 marks)

The absence of back-propagation would have made computation more difficult and complex which would lead to abandonment of some techniques.

Backward propagation hence made it more viable in feature creation from similar pieces of work.

4. If we never had a gradient descent approach, which other alternative technique can be used and why do we not commonly use it? (2 marks)

- Surrogate Optimization that uses a surrogate model to approximate the objective function which can be used when the objective function is expensive to evaluate
- Simulated Annealing: This can be defined as a probabilistic technique of approximating the global optimum of a function.
  The reason Simulated Annealing is not commonly used is because it allocated computation time is much more as compared to gradient descent.

5. Give two reasons why Apriori Algorithm is preferred to mining Association Rules (1 mark)

Apriori algorithm is a classic algorithm which works on a set of data in the database and provides us with the set of most frequent itemsets. It is used to mine the most frequent itemsets in a set of transactions.

Some of reasons why it is preferred are :

- The algorithm is exhaustive: This implies that it finds rules with adequate support.
- It provides an optimized sense of performance for all types of itemsets.

6. A new ML engineer suggests that we build a self-driving car "Visual System" using the Logistic Regression Model. Explain to him why this is not really a good idea (2 marks)

Models that are applied in Self-driving cars deal with a wide range of probabilities which include different environments and conditions to their automation.

A disadvantage is its high reliance on a proper presentation of the data. Logistic regression is not a useful tool unless all the important variables have been identified.

7. If we stick with Logistic Regression for this task, explain in detail three major changes we will have to make to the model/training-data (3 marks)

Normalization:- Featured can be scaled and normalized before being placed in the model.

Regularization: - This change will help with overfitting hence minimising loss.

Hyper- Parameter tuning: This will help in the accuracy of the model