DEEP LEARNING ASSIGNMENT 1

Professor: Chang - Hwan Lee

Sai Supraja Chinthapalli

Z number: Z23760554

DEEP LEARNING ASSIGNMENT: 1

ANSWER 1(i)(ii): In written pdf

ANSWER 2(i)(ii): In written pdf

ANSWER 2(iii):

To Draw Attention to Bigger Errors: Greater errors are amplified when the faults are squared. The error function penalizes larger differences between the actual and predicted values more severely. This is critical because minimizing the impact of huge errors is something we frequently want to focus because they can have a big influence on the model's overall performance.

To Guarantee Positivity: Squaring guarantees the positivity of each and every error. This is beneficial because it keeps negative and positive errors from canceling each other out when determining the overall error. The overall error is always guaranteed to be non-negative by summing the squared errors, which facilitates mathematical interpretation and manipulation.

To summarize, the error function's residuals are squared to highlight significant errors, maintain positivity, make optimization easier, and comply with statistical presumptions. Squared errors are a popular and useful option for calculating the difference between expected and actual values in regression situations because of these characteristics.

ANSWER 3:

Two primary issues can occur when applying linear regression to a classification problem where the target variable is categorical (e.g., binary or multi-class):

- **Breach of Assumptions**: Linear regression makes the assumption that there is a linear relationship between the independent and target variables. The link between features and class labels, however, can not always be linear in classification issues. When employing linear regression, the decision boundaries between classes are frequently non-linear, which results in subpar model performance. It is possible to get biased predictions and large error rates when trying to fit a linear model to non-linear data.

- **Non Probabilistic Predictions**: The results of linear regression are not probabilistic. For example, in binary classification tasks where class membership criteria are determined by probabilities, it is frequently necessary to collect probability estimates for each class in order to make well-informed decisions. Probabilities cannot be obtained directly using linear regression, and interpreting the continuous outputs as probabilities might be inaccurate and misleading.

ANSWER 4: In written pdf

ANSWER 5(i):

Given definition of linear classifier is

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w}^T \cdot \mathbf{x} > \theta, \\ 0 & \text{otherwise} \end{cases}$$

The above expression represents a binary linear classifier where:

f(x) is the predicted class (1 or 0)

w is the weight vector

x is the input vector

w^T .x is the dot product of weight vector and the input features

θ is the threshold

The decision boundary for logistic regression classifier is determined by the expression w^T .x> θ . If w^T .x> θ the predicted class is 1, otherwise it is 0.This decision rule is a characteristic of a linear classifier.In logistic regression the linear combination w^Tx is transformed using sigmoid function to obtain probability, but the decision boundary itself is linear with respect to the input function.

Therefore based on the given formula logistic regression can be considered as a linear classifier.This decision is made by comparing the linear combination w^Tx to the threshold.

ANSWER 5(ii):

In logistic regression the decision boundary is determined by the equation:

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

Here w is the weight vector

x is the input vector

b is the bias term

The decision boundary is formed when $w^Tx+b=0$

Solving the equation for $w^Tx+b=0$ gives the value of threshold which is similar to θ in the linear classifier definition. Therefore the threshold θ in logistic regression is given by -b.

So in logistic regression the value of θ is -b

ANSWER 6: In the written pdf

ANSWER 7(i)(ii): In written pdf

ANSWER 7(iii):

For classification and logistic regression, cross-entropy also referred to as log loss, is frequently favored over the sum of squared errors. Using cross-entropy has the following benefits:

- **Better in cases of classification issues:** Cross-entropy is especially meant for classification issues, in which predicting the likelihood of several classes is the main objective. Compared to the sum of squared errors, it penalizes inaccurate forecasts more severely and discrete outcome problems may not be a good fit for the sum of squared error.

- **Prevents the vanishing gradient issue:** During the training phase, the model parameters are updated using the gradient of the error concerning the weights.

The vanishing gradient issue, which might arise when using the sum of squared error, is mitigated by cross-entropy. In logistic regression, the derivative of the sigmoid function can result in very small gradients, which might impede learning, particularly in deep networks.

- **Gradient Descent Convergence:** When compared to sum of squared error, cross-entropy frequently results in faster convergence during gradient descent. This is especially crucial for deep learning models or high-dimensional spaces where training might be computationally costly. This is because the gradients with regard to the parameters are more illuminating and have a greater ability to direct the optimization procedure.

- **Prevents Learning Slowdown:** When the model is still a long way from the ideal answer in the early phases of learning, the gradients in the cross-entropy loss are bigger. This keeps the learning slowdown that can happen with squared error at bay. By doing this, the model may be able to learn faster and stay out of local minima.

- **Probabilistic Interpretation:** Information theory provides the basis for the probabilistic interpretation of cross-entropy. Since it calculates the difference between the actual and expected distributions, it fits better into probabilistic models like logistic regression, whose output is a probability distribution.

ANSWER 8: In written pdf

ANSWER 9: In boston housing jupyter file

ANSWER 10: In realestate valuation jupyter file