2.2

Friday, March 1, 2024 11:10 PM

1. a) Technically can be used for both (as seen in lecture) but better for regression
   b) Assumes a Gaussian distribution of errors in the output
   c) Advantage: Differentiable, can be used for Gradient descent
   Disadvantage: Sensitive to outliers, outliers may influence training too much
   The squared loss function imposes equal importance of all data and the heavy penalization of all outliers

2 a) Again, technically can be used for both but better for regression
   b) Does not make any assumptions about the output
   c) Advantage: Less sensitive to outliers than Squared Loss, more robust
   Disadvantage: Less efficient for optimization problems, gradient converges slower than Squared Loss
   Like Squared loss, Mean absolute error imposes equal importance of all data but with outliers not penalized as much

3 a) Classification
   b) Assumes output data represents probability
   c) Advantage: Penalizes incorrect predictions heavily, gradient is not 0 for confident wrong predictions
   Disadvantage: Limited application to classification, it relies on the assumption that the model is trying to predict probability for discrete classes
   Binary Cross-Entropy error imposes a probabilistic representation of classification on the model true labels of the data

4 a) Regression
   b) It assumes the presence of outliers and that outliers should not influence the model as much as other datapoints
   c) Advantage: Balanced sensitivity to outliers, provides a compromise between mean squared and absolute error
   Disadvantage: Introduces another hyperparameter that takes time to be chosen (δ-threshold)
   The Huber error imposes robustness to outliers and balanced sensitivity between all data points