Question 1: What are the different aspects you should consider when choosing a classification loss function? Please define each of these factors.

Question 2: Please describe the ‘Square loss function’ and the ‘Absolute loss function’ and the advantages/disadvantages of both.

Question 3: What are some difficulties in facial recognition and how do deep learning methods help solve these issues?

Reference:
https://github.com/spring-2023/csc-487-student/blob/main/materials/Survey_of_Loss_Functions.pdf

Answer 1: When choosing a classification loss function, the following factors should be considered: convexity, robustness and generalization performance. The convexity of a loss function refers to the shape of the graph created from the output of the loss function. The robustness of a loss function refers to the loss function's ability to handle outliers and noise within the data set. A loss function is more robust if it is able to minimize the impact of outliers/noise on the optimization of the algorithm. Generalization performance refers to a model's ability to perform on a test dataset that has not been seen yet.

Answer 2: The 'Square loss function' measures the square of the error between the true value y and the predicted value f(x) of the sample x, according to 'Survery of Loss Functions'. One disadvantage of this function is that because the error is squared, outliers are more noticeable and overall prediction performance with reduce as a result. One advtange of 'Square loss' is that the gradient is variable and is proportional to loss. This leads to fast convergence and high accuracy. 'Absolute loss function' measure the absolute value of the error between the true value y and the predicted value f(x) of the sample x. Contrary to 'Square loss', the gradient of 'Absolute loss' is fixed and this will negatively affect the efficiency of model training. 'Absolute loss' is more robust than 'Square loss' when there are outliers in the training set due to its fixed gradient. '

Answer 3: In facial recognition, each 'class' represents a single person. But for every class/person, these is unlikely to be a large amount of samples for that class compared to a class like 'dogs' or 'cats'. Deep learning extracts key information such as facial features that can be generalized and used to identify different classes/people. There are two main categories of loss functions in the facial recognition deep learning space: those that measure the difference between samples based on Euclidean space and those that measure the difference between samples in the angular space.