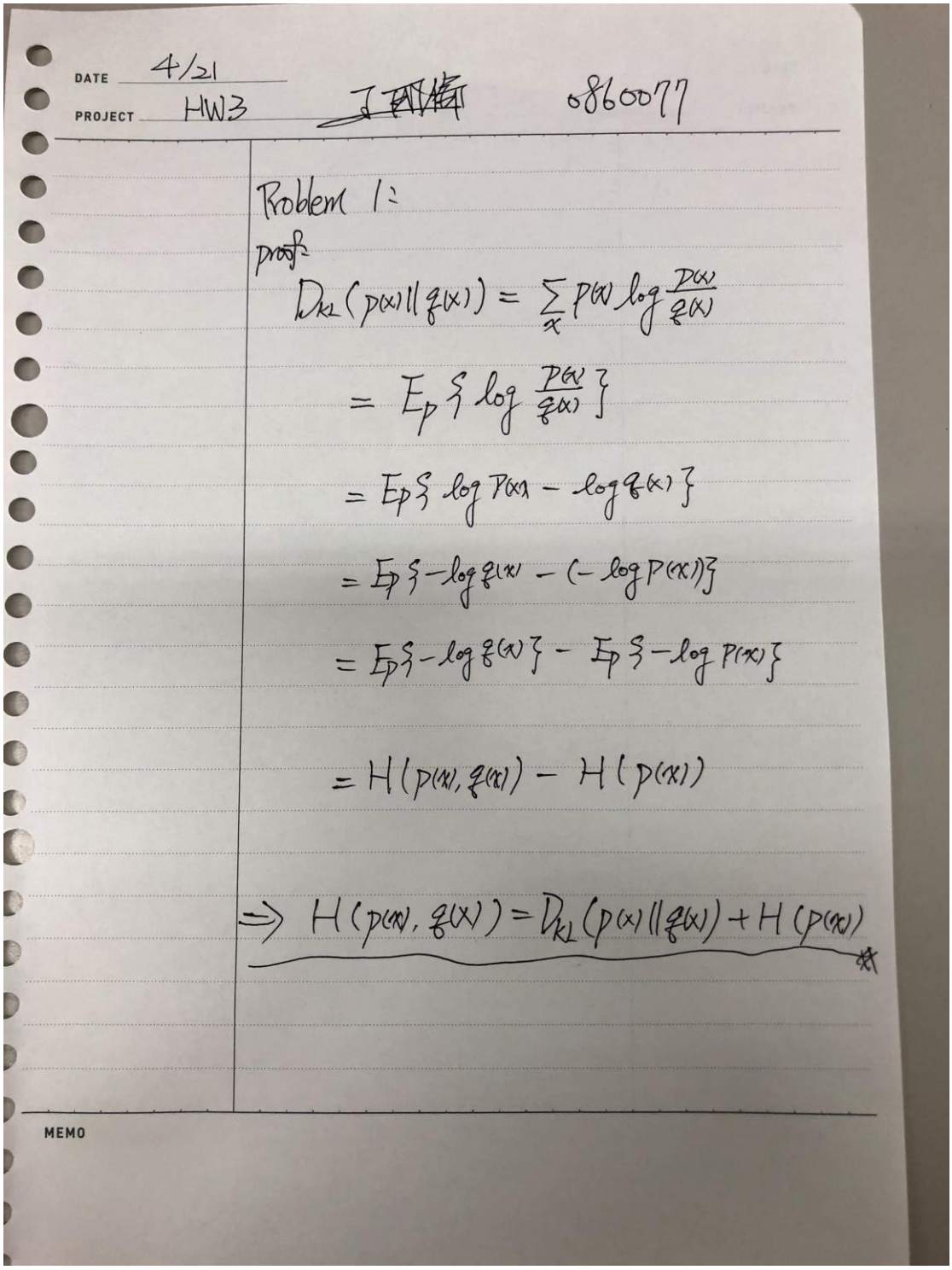
姓名:王國倫　　學號:0860077　　Homework3

Problem 1:



Problem 2:

The detailed code in HW3_0860077_王國倫.ipynb

Problem 3:

(a)MSE

Advantage:

This loss function can ensure that our trained model doesn't exist huge outlier, because of square part of the function, it will penalize outlier with the large errors.

Disadvantage:

In generally, our goal is training a well-rounded model that meets majority cases, so we don't care much about the outlier case, these situations have a small probability of occurrence. However, the square part of the function will calculate a huge error, and match these outliers, leading to bad performance.

(b)MLE

Advantage:

The MLE improve the MSE disadvantage, this loss function gives every error the same weight, so we won't be putting too much weight on these outliers, and provides a general prediction meets the most situations.

Disadvantage:

If all case has an equal probability of occurrence, the trained model by using the MLE loss function, this might result in our model have a poor performance, that's because we don’t take these outliers into training model, the loss function must be give appropriately weight to every error.

(c)Hinge Loss for SVM

According to this loss function, it will calculate all incorrect class, and obtain different of between with right and wrong case, take the large of the value. After plus threshold, and compare with zero, select the bigger value as the loss of this class. The maximum and minimum of Hinge loss are zero and infinite, respectively. This also points out that only when the classification is completely correct, there will be zero loss. As a result, this loss function not only get correct class, but also ensure that the score of correct class higher than the incorrect class.

(d)Cross Entropy Loss

Advantage:

When using gradient descent renew parameter, the speed of convergence depends on the learning rate and gradient, in particularly, the gradient was calculated from the loss function. So, if model has bad performance, the speed of learning is faster, in contrary, the model has great performance, the speed of learning is slow.

Disadvantage:

Cross Entropy Loss only care about the accuracy of the prediction probability of the correct label, ignoring the difference of other incorrect labels, resulting features learned discretely.

problem 4:

The both of methods have same regularization term L2, deal with the model easy to overfitting, and have a bigger tolerance for outlier, decreasing the opportunity of overfitting, provider a general prediction.