Programming Assignment 2 Report

ChenXiao Liu

Introduction

Problem & Motivation

The problem is to recognize letters from image data. There are a lot of classifiers could solve this problem, but different classifiers have different accuracy and efficiency, which is the motivation for measuring the performance of those classifiers and determining the best classifier on letter image recognition data. To determine the best classifier, measurements such as the validation accuracy and the time cost would be used.

Binary classification problem introduction

Binary classification refers to classify items in a set into two groups based on some specific classification methods. I chose letter A and B for the third problem. Letter O and I might be the easiest pair to classify because they are very distinct from each other; letter O and Q might be the hardest pair to classify, because they look like very similar.

Why dimension reduction is or is not useful for this problem

Normally dimension reduction is useful because it could reduce computational cost and the number of samples required to fit a model, and some features might not be important. It's hard to say which dimension reduction methods are better. If a dimension reduction method is good depends on the dataset and the model prediction accuracy. Therefore, if a dimension reduction method could greatly reduce the model fit time and the computational cost, or improve the validation accuracy, then it's a good dimension reduction method.

Results

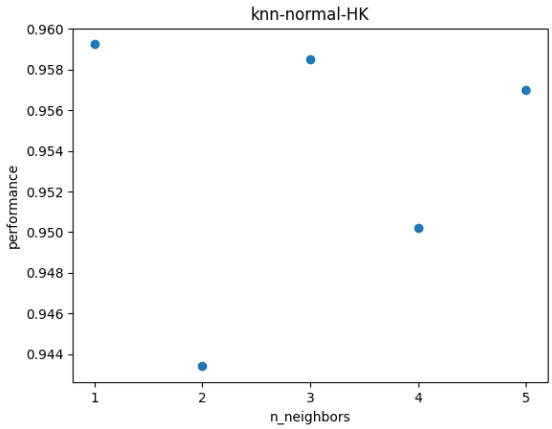
k-nearest neighbors

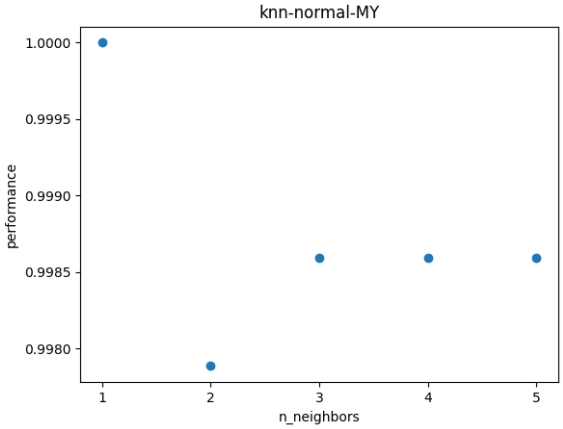
Description

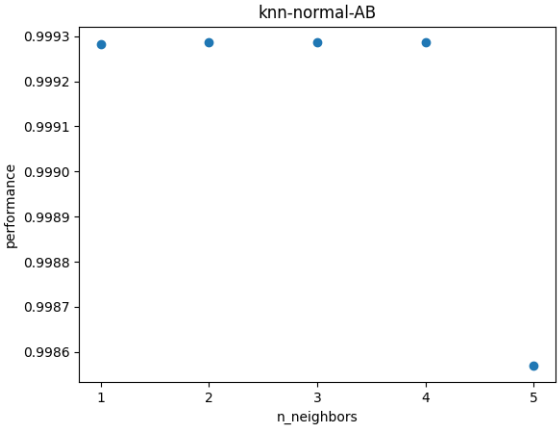
k-nearest neighbors is a commonly used classification method. It classifies a sample by choosing labels of the k samples whose response values are most similar. The advantage of k-nearest neighbors is that it does not require to train a model and it's easy to explain. The disadvantage is that it might take too much time to compute distances and it might require store great amount of data.

Hyperparameter: k

Cross validation results



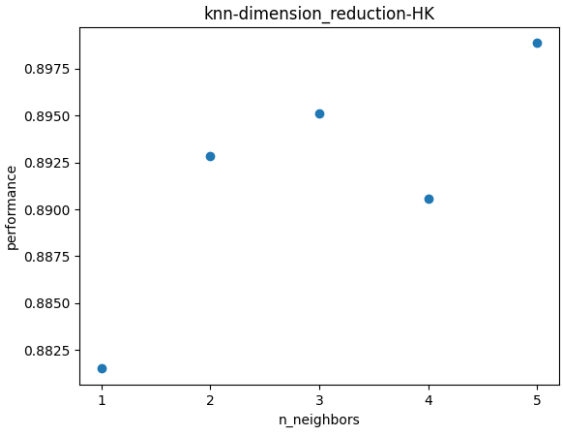


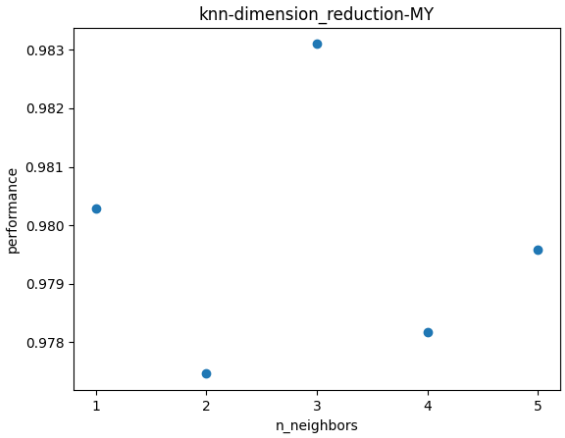


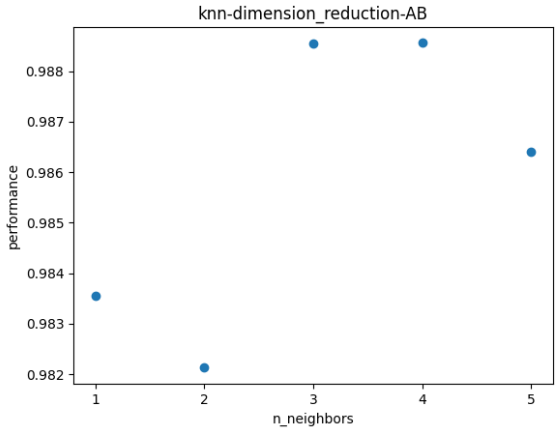
Dimension reduction method

Feature extraction - PCA

Cross validation results with dimension reduction







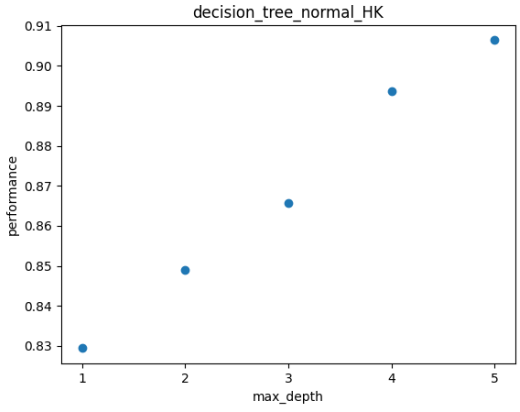
Decision Tree

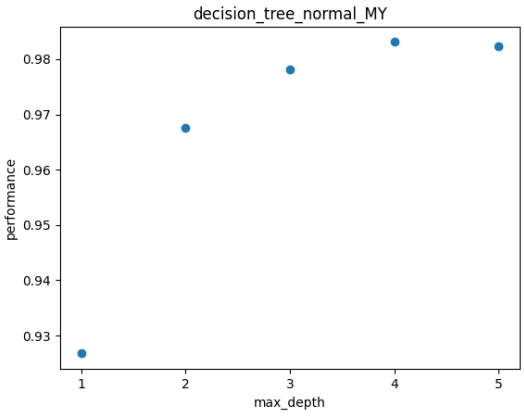
Description

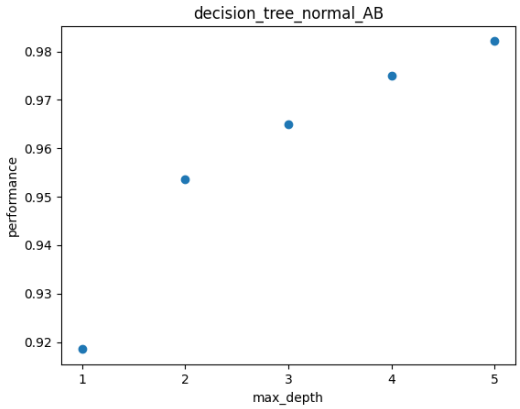
Decision tree utilizes a model that looks like a tree to make decisions. The advantage of a decision tree is that it is easy to apply, it's splits could be easily understood. The disadvantage is that it would cause overfit problem.

Hyperparameter: max depth

Cross validation results



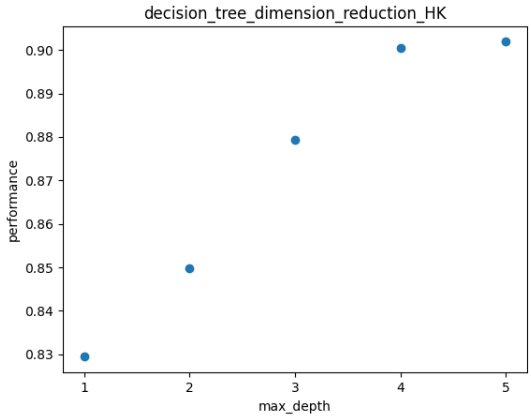


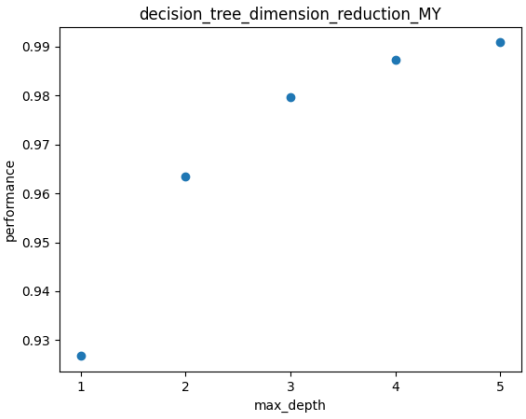


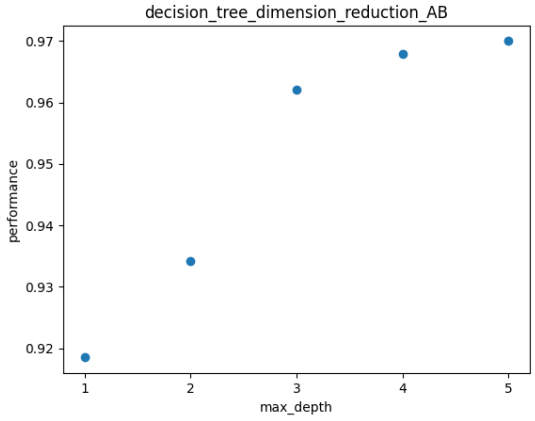
Dimension reduction method

Filter Method – ANOVA (Analysis of Variance)

Cross validation results with dimension reduction







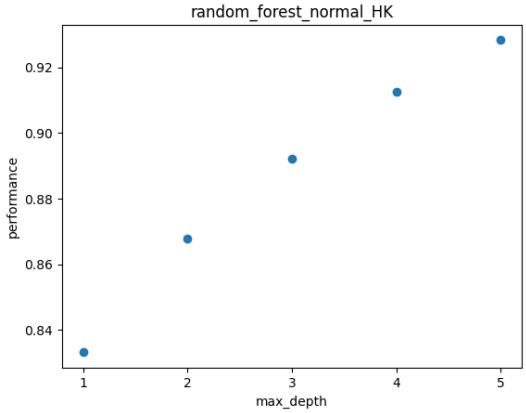
Random Forest

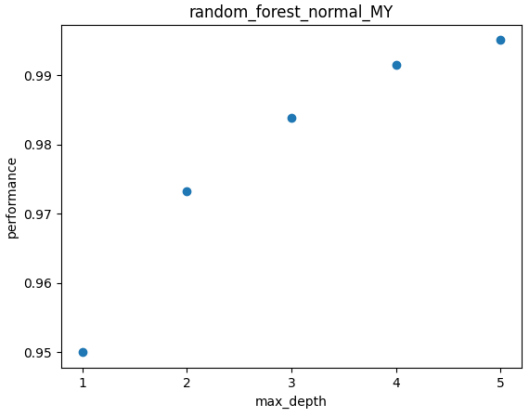
Description

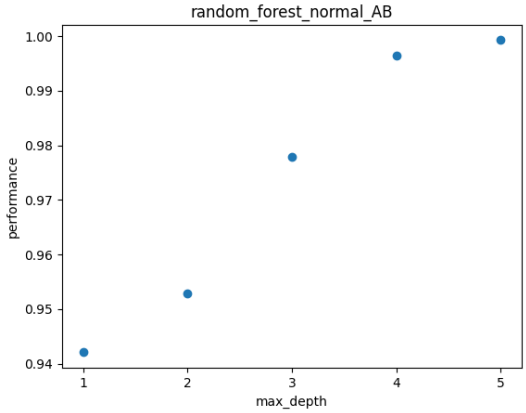
Random forest is an ensemble learning method, which combines multiple decision trees to solve a problem. The advantage of a random forest is that it's very accurate and could avoid overfit. The disadvantage is that it might cost too much time to train.

Hyperparameter: max depth

Cross validation results



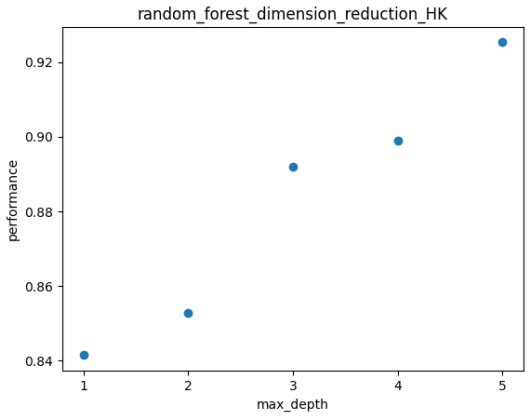


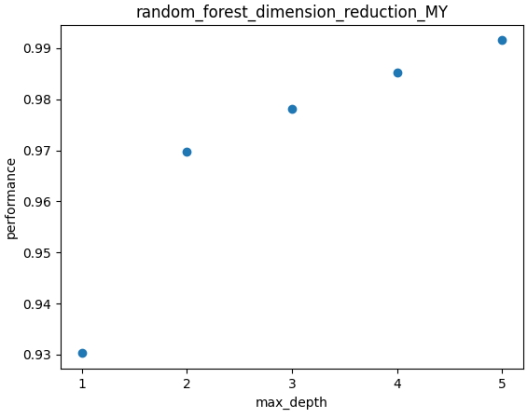


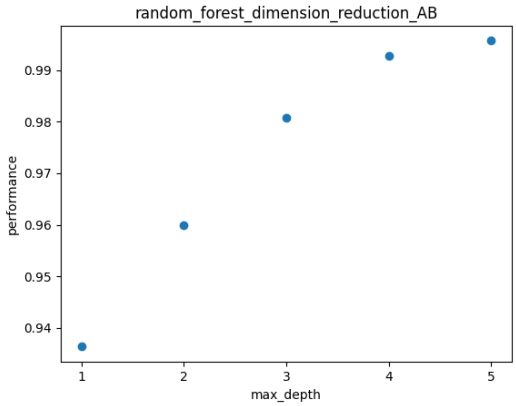
Dimension reduction method

Wrapper feature selection - Greedy Backward Feature Elimination

Cross validation results with dimension reduction







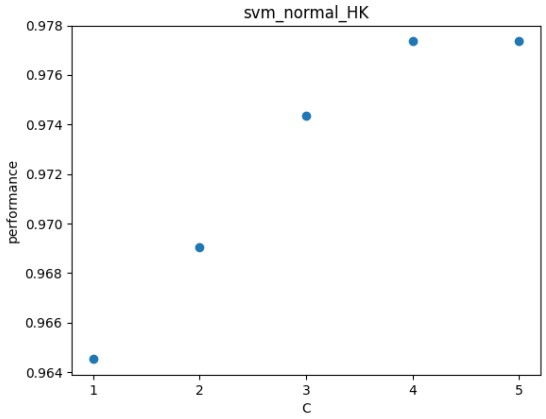
SVM

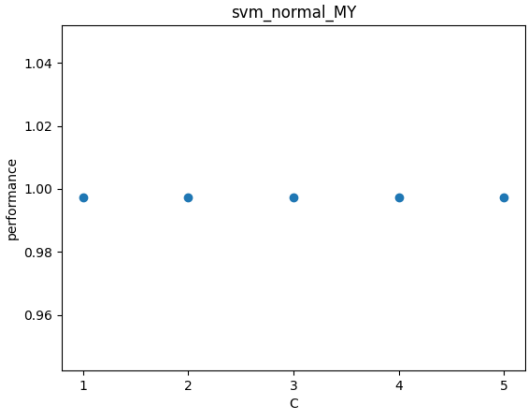
Description

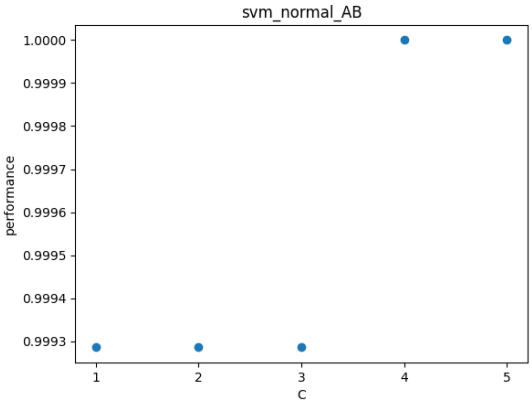
SVM (support-vector machine) is a supervised learning model that could be applied to classification and regression problems. The advantage of SVM is that it's very efficient in high dimension and nonlinear classification. The disadvantage is that if the dataset contains too many meaningless samples, then the performance of SVM might be bad.

Hyperparameter: C

Cross validation results



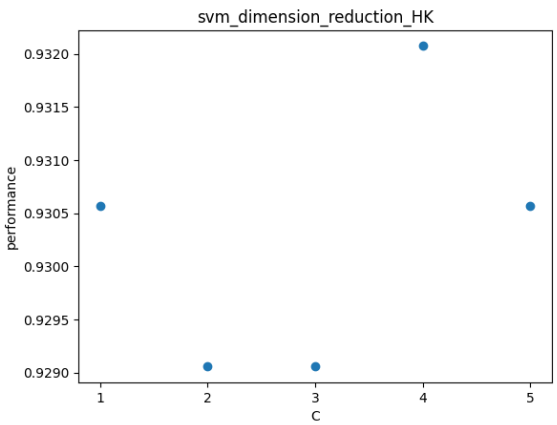


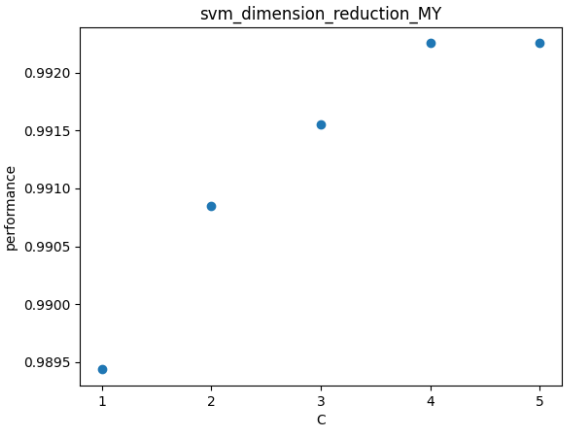


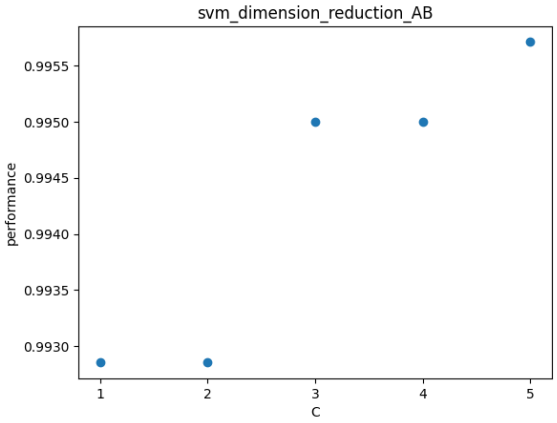
Dimension reduction method

Wrapper feature selection - Greedy Forward Feature Construction

Cross validation results with dimension reduction







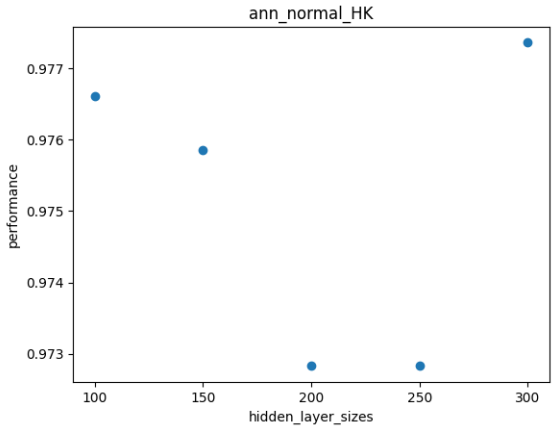
Artificial Neural Network

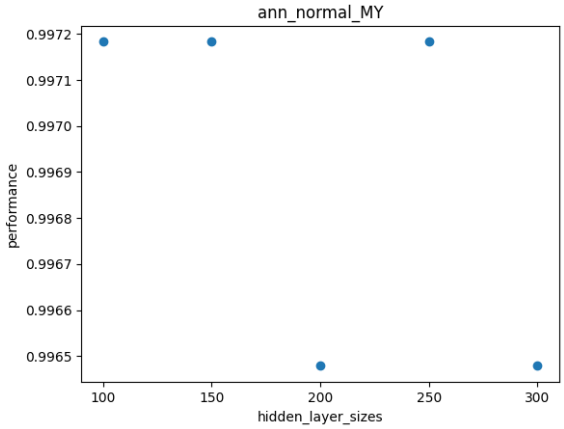
Description

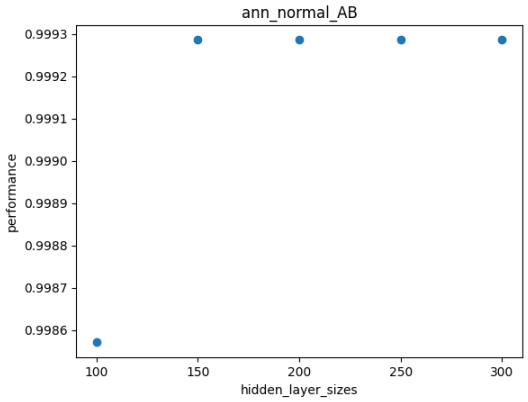
An Artificial Neural Network (ANN) is a model consists of layers of connected nodes. The advantage of an ANN is that it's accurate and could fit any models. The disadvantage is that it's difficult to understand and explain.

Hyperparameter: hidden layer sizes

Cross validation results



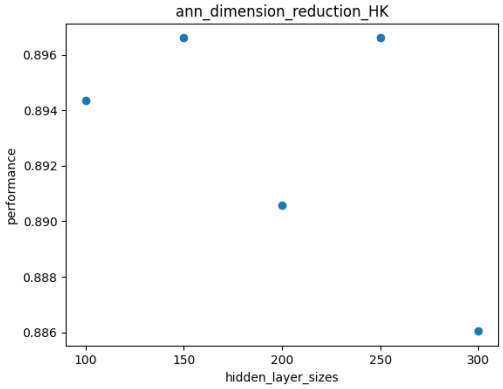


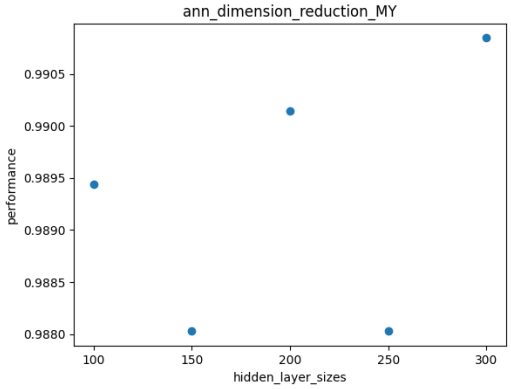


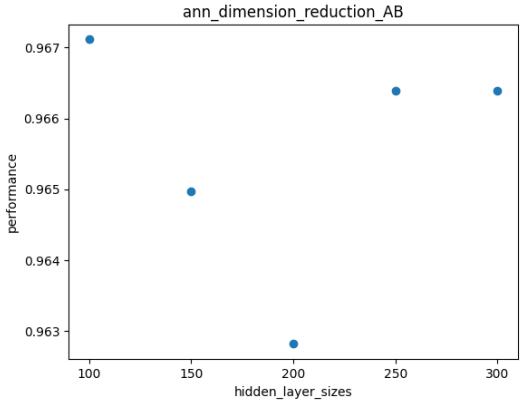
Dimension reduction method

Filter Method – Chi-Square Test

Cross validation results with dimension reduction







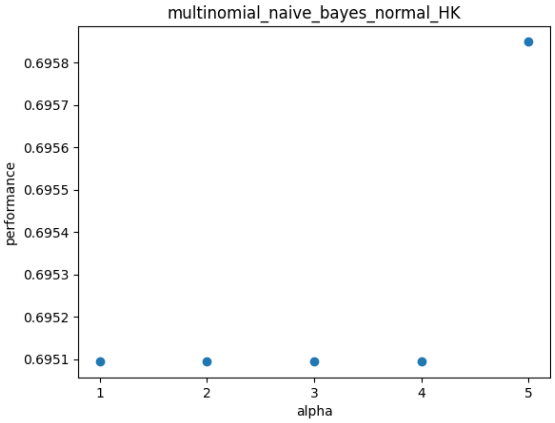
Multinomial Naive Bayes

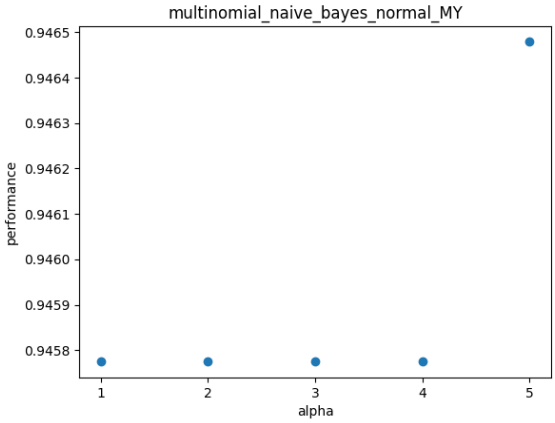
Description

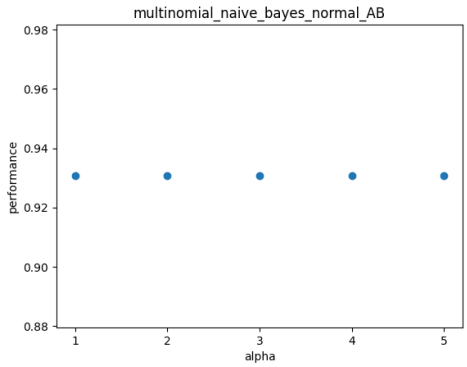
Multinomial Naive Bayes is a supervised learning approach used for text classification. The advantage of it is that it only needs to compute probability, so it's easy to implement. The disadvantage is that it not accurate.

Hyperparameter: alpha

Cross validation results



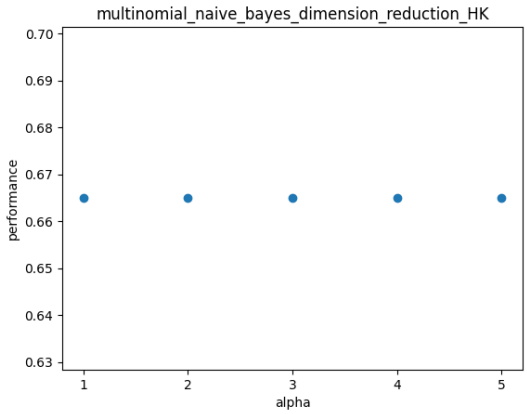


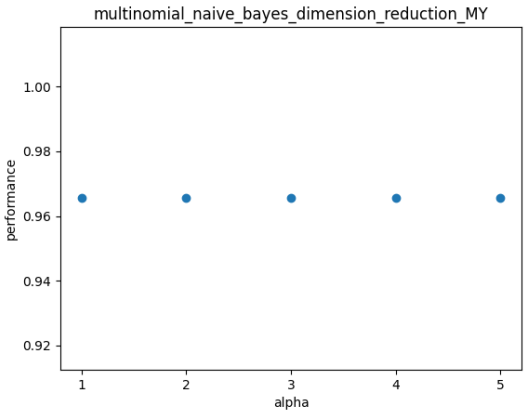


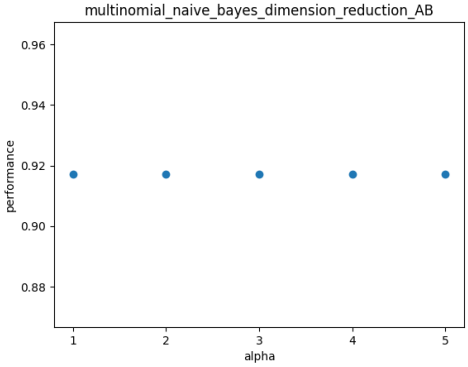
Dimension reduction method

Wrapper Feature Selection - Recursive feature elimination

Cross validation results with dimension reduction







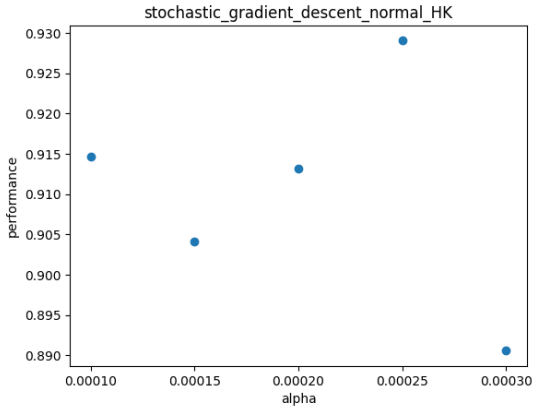
Stochastic Gradient Descent

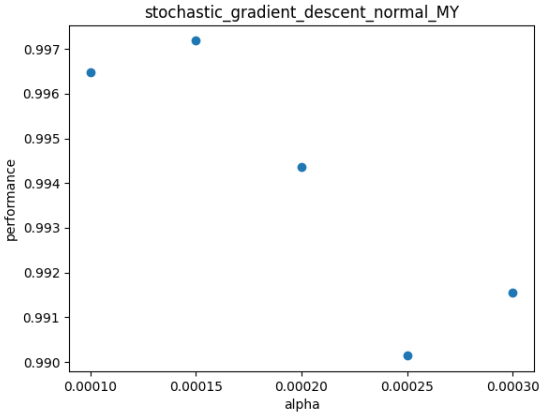
Description

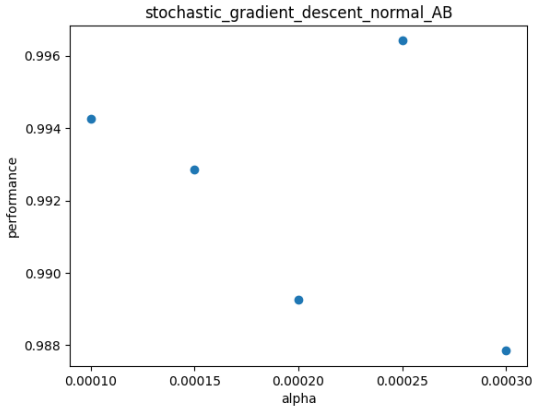
Stochastic gradient descent is used for optimizing an objective function iteratively. The advantage of it is that it computes fast during each iteration. The disadvantage is that it might take too much time to find the minima of the objective function.

Hyperparameter: alpha

Cross validation results



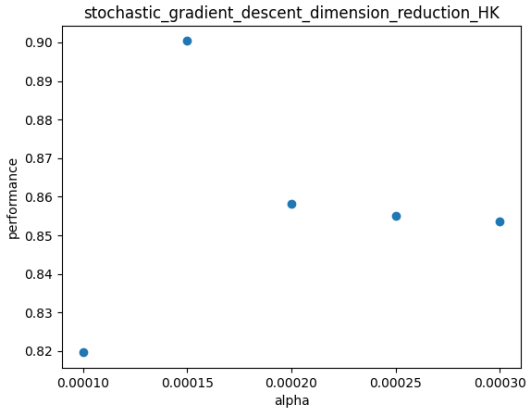


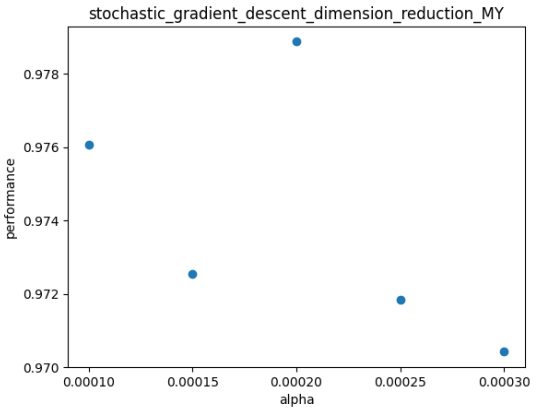


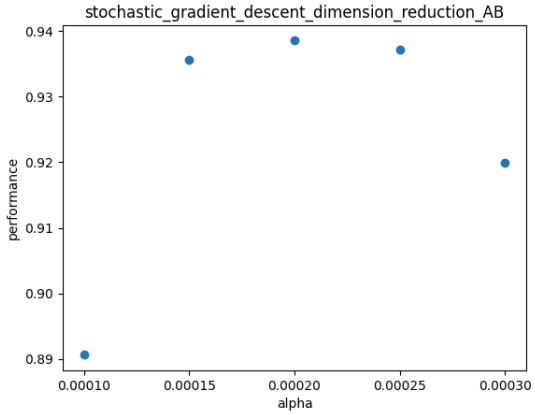
Dimension reduction method

Filter Method – ANOVA (Analysis of Variance)

Cross validation results with dimension reduction







Discussion

Comparison before dimension reduction

Classify Pair H & K before dimension reduction		
Model	Score (Performance)	Runtime (Seconds)
KNN	0.9391891891891891	0.010026931762695312
Decision Tree	0.8986486486486487	0.0003230571746826172
Random Forest	0.9054054054054054	0.004121065139770508
SVM	0.9662162162162162	0.0016438961029052734
ANN	0.972972972972973	0.00592803955078125
Multinomial Naive Bayes	0.6351351351351351	0.0003120899200439453
Stochastic Gradient Descent	0.9121621621621622	0.00031065940856933594

Classify Pair M & Y before dimension reduction		
Model	Score (Performance)	Runtime (Seconds)
KNN	1.0	0.018830060958862305
Decision Tree	1.0	0.0003247261047363281
Random Forest	0.9936708860759493	0.003720998764038086
SVM	1.0	0.0008111000061035156
ANN	1.0	0.009374141693115234
Multinomial Naive Bayes	0.9556962025316456	0.0002789497375488281
Stochastic Gradient Descent	0.9873417721518988	0.00031113624572753906

Classify Pair A & B before dimension reduction		
Model	Score (Performance)	Runtime (Seconds)
KNN	1.0	0.005321025848388672
Decision Tree	0.9871794871794872	0.0002949237823486328
Random Forest	1.0	0.0037376880645751953
SVM	1.0	0.0006647109985351562

ANN	0.9871794871794872	0.00047206878662109375
Multinomial Naive Bayes	0.9423076923076923	0.0002999305725097656
Stochastic Gradient Descent	0.9871794871794872	0.0002911090850830078

Comparison after dimension reduction

Classify Pair H & K after dimension reduction		
Model	Score (Performance)	Runtime (Seconds)
KNN	0.8918918918918919	0.0019059181213378906
Decision Tree	0.8513513513513513	0.00027489662170410156
Random Forest	0.8918918918918919	0.003785371780395508
SVM	0.9324324324324325	0.0017561912536621094
ANN	0.8851351351351351	0.0003190040588378906
Multinomial Naive Bayes	0.6081081081081081	0.00028967857360839844
Stochastic Gradient Descent	0.8378378378378378	0.00024700164794921875

Classify Pair M & Y after dimension reduction		
Model	Score (Performance)	Runtime (Seconds)
KNN	0.9936708860759493	0.0019249916076660156
Decision Tree	0.9936708860759493	0.000263214111328125
Random Forest	0.9810126582278481	0.004015922546386719
SVM	0.9936708860759493	0.0007047653198242188
ANN	0.9936708860759493	0.004672050476074219
Multinomial Naive Bayes	0.9936708860759493	0.00027489662170410156
Stochastic Gradient Descent	0.9810126582278481	0.0002498626708984375

Classify Pair A & B after dimension reduction		
Model	Score (Performance)	Runtime (Seconds)
KNN	0.9551282051282052	0.0022668838500976562
Decision Tree	0.9615384615384616	0.0002582073211669922
Random Forest	1.0	0.0036499500274658203
SVM	0.9935897435897436	0.0007450580596923828
ANN	0.9551282051282052	0.002180814743041992
Multinomial Naive Bayes	0.8846153846153846	0.0002391338348388672
Stochastic Gradient Descent	0.9102564102564102	0.00024890899658203125

Lessons learned

I would choose SVM for this letter classification problem, because the average performance of SVM is the best and the average runtime is good as well. The dimension reduction decreases the accuracy, but greatly reduces the runtime, I guess that's because some meaningful features have been deleted, therefore the accuracy decreased; and the computational cost required for 4 features is less than that of 16 features, so the runtime decreased as well. If given a new dataset, I would not perform dimension reduction because I think the accuracy is much more

important and I can accept the task require more time without dimension reduction. This project made me think about that choose a correct model is important; the selection of model depends on the criterion, and dimension reduction is not suitable for all tasks; it's ok to delete more features if the runtime is more important, but if the accuracy is more important, only meaningless features should be deleted.