Classification

In this section, we implement multi-class classification models on the Fashion-MNIST dataset. We first use two algorithms, k-Nearest Neighbor (k-NN) and Linear Discriminant Analysis (LDA), which can be directly used for multiclass classification. Then, we extend Support Vector Machine (SVM), which is a binary classifier, to handle the multiclass case.

k-Nearest Neighbors

We first implement k-NN classification model to classify the Fashion-MNIST data. The Euclidean distance is used to measure the distance between variables. The raw data are scaled in the data pre-processing step, so the distance is scale-invariant.

To determine the parameter k, we use 5-fold cross-validation and calculate the overall mis-classification rate for each k. The figure shows the relation between mis-classification rate and k. We find that the best mis-classification rate is reached at $k = 10$.

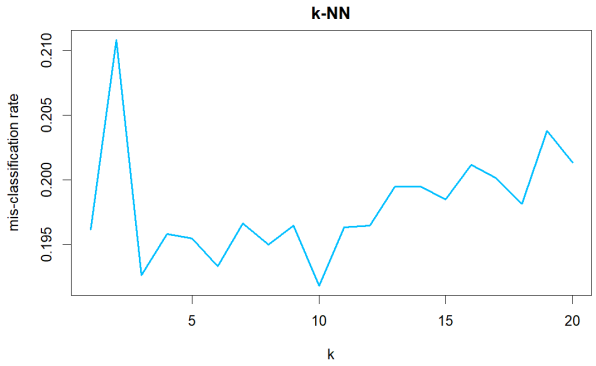

Figure 1: 5-fold CV Mis-classification Rate

Based on the result of cross-valiadation, we choose $k = 10$ to fit the k-NN model. The confusion matrix and mis-classification rate for each class are shown in the following two tables. For Fashion-MNIST data, the overall mis-classification rate of the 10-NN classifier is 0.144 and the accuracy is 0.856.

Ytest	Ytest_pred 0	1	2	3	4	5	6	7	8	9
0	856	1	18	17	4	0	95	1	8	0
1	3	964	9	13	1	0	10	0	0	0
2	11	0	794	11	97	0	86	0	1	0
3	30	8	13	888	37	0	24	0	0	0
4	2	0	74	22	810	0	90	0	2	0
5	0	0	0	1	0	788	10	121	6	74
6	191	1	106	19	71	0	601	0	11	0
7	0	0	0	0	0	2	0	936	0	62
8	2	1	15	3	5	0	11	6	955	2
9	0	0	0	0	0	1	1	31	0	967

(a) Confusion Matrix

class	mis−classification
0	0.218
1	0.011
2	0.228
3	0.088
4	0.21
5	0.004
6	0.352
7	0.145
8	0.028
9	0.125
overall	0.144

(b) Mis-classification Rate

Figure 2: 10-NN

LDA

We then classify the Fashion-MNIST data using the LDA classifier. The confusion matrix and mis-classification rate for each class are shown in the following two tables. The overall mis-classification rate of the LDA classifier is 0.174 and the accuracy is 0.826. We find that the accuracy of LDA classifier is slightly lower than that of 10-kNN classifer.

	Ytest_pred									
Ytest	0	1	2	3	4	5	6	7	8	9
0	863	0	12	25	0	2	88	0	10	0
1	2	976	1	15	0	1	5	0	0	0
2	10	0	835	14	78	0	59	0	4	0
3	24	5	7	923	22	0	18	0	1	0
4	0	1	57	26	870	0	44	0	2	0
5	0	0	0	0	0	947	0	35	4	14
6	148	0	74	24	57	0	689	0	8	0
7	0	0	0	0	0	15	0	952	0	33
8	3	0	6	3	1	2	9	2	974	0
9	0	0	0	0	0	8	0	33	0	959

(a) Confusion Matrix

class	mis−classification
0	0.198
1	0.005
2	0.244
3	0.178
4	0.254
5	0.138
6	0.414
7	0.128
8	0.062
9	0.103
overall	0.174

(b) Mis-classification Rate

Figure 3: LDA

SVM

We have applied two multiclass classifiers k-NN and LDA for the Fashion-MNIST classification. Now, we extend SVM to the multiclass form to solve the classification problem. SVM is a binary classifier and it does not support multiclass classification natively. However, we can break the multiclass classification problem into several binary ones. There are two common methods to extend SVM for multiclass classification, One-vs-One approach and One-vs-Rest approach.

In One-vs-One approach, we fit SVM models for every two classes. Each classifier separates points of two different classes. Suppose we have k classes, we then fit $\frac{k(k-1)}{2}$ SVM models. In prediction stage, we input data into all binary classifiers. Each binary classifiers will decide a class that the input is belonged to. We let those binary classifiers vote for the class of input, and the prediction result is the class that most classifiers vote for.

In One-vs-Rest approach, we fit SVM models to distinguish points of one certain class from the other classes. Suppose we have k classes, we then fit k SVM models. In prediction stage, we input data into all One-vs-Rest classifiers. Each classifier will give a probability that the input is belonged to that class. The prediction result is the class with the largest probability.

One-vs-Rest approach is more computationally efficient than One-vs-One approach, since it only need to fit k SVM models while One-vs-One approach need to fit $\frac{k(k-1)}{2}$ models. However, in One-vs-Rest approach, the training data is unbalanced since the ratio of training data from each class is $1 : (k-1)$, which may cause biase.

We fit the multiclass SVM model using One-vs-One approach, with using radial basis kernel. The overall mis-classification rate of kernel SVM is 0.09 and the accuracy is 0.91. The kernel SVM classifier outperforms k-NN and LDA.

In addition, a linear SVM model is constructed to compare with the RBF kernel. The overall mis-classification rate of SVM is 0.09 and the accuracy is 0.91.

	Ytest_pred_svm									
Ytest	0	1	2	3	4	5	6	7	8	9
0	879	0	15	17	1	1	81	0	6	0
1	2	987	0	7	0	1	3	0	0	0
2	17	0	826	17	76	0	61	0	3	0
3	26	10	8	918	25	0	12	0	1	0
4	3	1	59	24	872	0	39	0	2	0
5	0	0	0	0	0	959	1	27	3	10
6	121	0	53	28	49	0	745	0	4	0
7	0	0	0	0	0	10	0	962	0	28
8	2	0	4	2	1	2	5	1	983	0
9	0	0	0	0	0	2	0	26	0	972

(a) Confusion Matrix

class	mis−classification
0	0.163
1	0.011
2	0.144
3	0.094
4	0.148
5	0.016
6	0.213
7	0.053
8	0.019
9	0.038
overall	0.09

(b) Mis-classification Rate

Figure 4: Kernel SVM

Conclusion

In this part, we build k-NN, LDA, linear SVM and kernel SVM to classify Fashion-MNIST data. The performace of different algorithms are shown in the table. The accuracy of four methods: Kernel SVM > k-NN > LDA.

Table 1: Model Summary

	kNN	LDA	Kernel.SVM
Accuracy	0.8559	0.8256	0.9103