# ECS 271 Homework1

### Jiahui Guan, Leyuan Wang

## 1 Problem 1

We implemented k-nearest neighbor algorithm with three different k's: k=1,k=5,k=50. When testing a data, we calculate its distances of all the training data. Note that here we use Euclidean distance and in this case its a 16-dimensional Euclidean distance. Then we choose the nearest k points. Nearest means the smallest k Euclidean distance from the specific testing data. Note that when k=1, we just only pick up the nearest point. After getting those nearest neighbors, we count which class they fall into. Here is which digits they "come from". We pick up the one that appears the most.

The algorithm is pretty easy to implement without using any package. We use R to write the code. Detailed R code can be found at Appendix. One thing needed to be noticed is that when k = 1, it will automatically result in overfitting. Similarly, when k is very large, it will produce underfitting. Here for k = 50, it's hard to say it will cause underfitting without looking into the real data.

We use Python to implement Support Vector Machine (SVM).

#### 2 Problem 2

Due to time limitation, we perform 3-fold cross validation for the training dataset. Below is the error rate for each fold and their total mean error

algorithm	error rate fold1	error rate fold2	error rate fold3	mean error
k=1	0.00886	0.01120	0.01040	1.0141%
k=5	0.0096	0.01120	0.02081	1.3877%
k=50	0.0904	0.08086	0.1	9.047%
linear kernel	8.006e-03	8.006e-03	8.807e-03	0.827%
polynomial (p=2)	5.60e-03	8.80e-03	1.36e-02	0.934%

Also, we can see the confusion matrix for each algorithm

• k-nn, when k=1

У predict0 129 0 128 0 117 0 138 0 134 0 131 0 125 0 131 0 111

• k-nn, when k=5

у preict 0 0 139 0 121 3 135 0 115 0 122 0 138 0 114 0 122 0 119 0 107

• k-nn, when k=50

у predict 0 0 134 19 129 0 116 0 130 

```
      5
      0
      0
      0
      0
      0
      110
      0
      0
      1
      10

      6
      3
      1
      0
      0
      0
      129
      0
      1
      0

      7
      0
      1
      1
      0
      0
      0
      0
      111
      3
      0

      8
      0
      0
      0
      0
      0
      0
      107
      0

      9
      0
      1
      0
      0
      1
      4
      0
      0
      0
      84
```

# 3 Problem 3

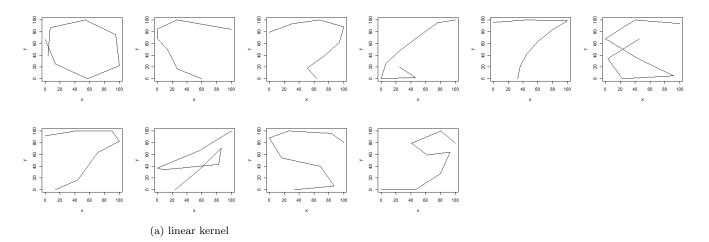


Figure 1

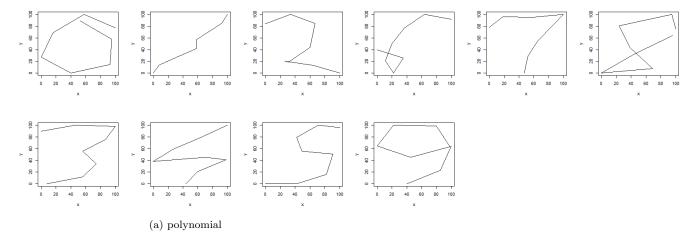


Figure 2

Above the is the example instance corresponding to the support vector for each digit using different kernels:linear, polynomial (p=2). We somehow can tell that it is handwriting from digit 0-9.

kernel	total number of support vectors	
linear	486	
polynomial(p=2)	981	

Just by looking at the number of support vectors, linear kernel might be have a better result.

# 4 Problem 4

KNN assumes that the data is in a feature space. More exactly, the data points are in a metric space. The data can be scalars or possibly even multidimensional vectors. Since the points are in feature space, they have a notion of distance This need not necessarily be Euclidean distance although it is the one commonly used.

# 5 Problem 5

• K-nn, when k=1

[3181] 2 2 0 3 6 6 4 8 4 0 1 7 2 3 6 9 2 4 6 5 5 8 3 1 5 0 2 8 4 2 5 9 7 1 0 0 8 4 9 3 3 3 2 8 3 0 6 9 2 9 6 4 3 [3234] 4 3 6 3 0 4 7 5 2 8 7 6 0 4 2 3 4 5 9 0 6 9 1 2 8 1 0 8 4 3 5 8 9 5 8 5 2 3 6 2 8 4 1 2 1 2 4 8 1 9 7 0 6 [3287] 8 6 7 6 7 0 9 7 3 3 2 6 0 6 7 6 3 0 2 7 2 3 8 5 6 8 4 0 7 7 2 8 0 5 6 9 5 2 2 9 1 9 2 3 0 4 1 5 2 0 1 4 0 [3340] 3 9 6 7 1 2 3 6 7 6 5 8 5 6 2 2 3 6 6 6 7 7 2 4 9 6 4 0 8 4 8 7 8 5 7 4 2 6 7 7 2 3 6 2 3 7 6 7 2 8 2 9 7 [3393] 2 3 0 7 2 1 9 0 3 7 6 0 8 1 5 2 9 5 3 9 8 5 2 9 2 5 1 6 8 3 2 8 9 7 0 1 2 0 0 8 3 5 7 0 9 9 7 9 1 4 1 2 5 [3446] 9 0 2 4 6 1 0 0 9 4 0 7 7 6 5 2 5 1 2 9 5 4 5 8 8 9 7 5 9 1 6 7 7 8 5 3 2 3 6 0 1 9 8 0 7 2 1 5 1 1 4 0 [3446] 9 7 7 4 9 5 5 1 4 4 3 3 9 2 7 7 1 4 0 3 5 6 0 8 4 0 8 5 6 5 7 3 0 4 1 5 7 3 2 6 6 7 7 1 2 2 6 9 0 7 8 1 0 4 [3552] 1 7 8 3 9 2 7 0 1 8 6 2 8 5 6 9 7 7 4 4 5 0 8 7 7 0 2 3 7 6 4 8 3 1 3 6 2 9 3 7 8 4 4 3 3 9 2 4 3 1 2 3 3 1 0 1 6 [3658] 3 4 6 7 3 4 5 8 2 7 6 0 9 6 0 1 7 1 8 9 8 7 1 5 4 4 8 1 4 1 6 6 0 3 5 7 2 7 5 4 1 2 0 1 9 0 2 6 4 8 9 3 3 [3711] 0 3 5 9 4 3 0 3 7 7 0 3 3 3 1 1 1 3 6 8 4 2 7 6 4 1 9 2 6 3 9 0 5 4 5 1 7

#### • K-nn, when k=5

#### • K-nn, when k=50

[3175] 8 4 8 7 0 0 2 2 0 3 6 6 4 8 4 0 1 7 2 3 6 9 2 4 6 5 5 8 3 1 5 0 2 7 4 2 5 9 7 6 0 0 8 4 9 3 [3221] 3 3 2 8 3 0 6 9 2 9 6 4 3 4 3 6 3 0 4 7 5 2 8 7 6 0 4 2 3 4 5 9 0 6 9 1 2 8 1 0 8 4 3 5 8 9 [3227] 5 8 8 5 2 3 6 2 8 4 1 2 1 2 4 8 1 9 7 0 6 8 6 7 6 7 4 9 7 3 3 2 6 0 6 7 6 3 0 2 7 2 3 8 5 6 8 5 [3313] 4 0 7 7 2 4 9 6 4 0 8 4 1 7 8 5 7 4 2 6 7 7 2 3 6 2 3 7 6 7 2 7 2 9 3 2 3 0 7 2 1 9 0 3 7 6 0 [3359] 6 7 7 2 4 9 6 4 0 8 4 1 7 8 5 7 4 2 6 7 7 2 3 6 2 3 7 6 7 2 7 2 9 3 2 3 0 7 2 1 9 0 3 7 6 0 [3455] 8 1 5 2 9 5 3 9 8 5 2 9 2 5 1 6 8 3 2 8 9 7 0 2 2 0 0 8 3 5 7 0 9 9 7 9 1 4 2 2 5 9 0 2 4 6 [3451] 1 0 0 9 4 0 7 7 6 5 2 5 1 2 9 5 4 5 8 8 9 7 5 9 2 6 7 7 7 8 5 3 2 3 6 0 1 9 8 0 7 2 1 5 1 1 [3497] 4 0 7 7 4 9 5 5 1 4 4 3 3 0 2 7 7 1 4 0 3 5 6 0 8 4 0 7 5 6 5 7 3 0 4 1 5 7 3 2 6 6 1 7 1 2 [3543] 2 6 9 0 7 8 2 0 4 1 7 8 3 9 2 7 0 1 8 6 2 8 5 6 0 8 4 0 7 5 6 5 7 3 0 4 1 5 7 3 2 6 6 1 7 1 2 [3543] 3 6 9 8 7 8 9 8 7 8 9 2 4 3 1 3 0 3 7 7 0 3 3 3 1 2 [3681] 5 4 4 8 1 4 1 6 6 0 3 5 7 2 7 5 4 1 2 0 1 9 0 2 6 4 8 9 3 3 0 3 5 9 4 3 0 3 7 7 0 3 3 3 1 2 [3681] 5 4 4 8 1 4 1 6 6 0 3 5 7 2 7 5 4 1 2 0 1 9 0 2 6 4 8 9 3 3 0 3 5 9 4 3 0 3 7 7 0 3 3 3 1 2 [3727] 2 3 6 7 4 2 7 6 4 1 9 1 6 3 9 0 5 4 5 2 7

- SVM, for linear Kernel
- SVM, for polynomial Kernel
- SVM, for Gaussian kernel

## 6 Problem 6

Note that by doing transfer learning, we need to add  $Y(W_s^TX + W_{os})$  (which is constant) to our constraint. That is,  $Y(W_s^TX + W_{os}) + Y(W_t^TX + W_{ot}) \ge 1$ .

Objective function:

$$\arg\min_{w_t, w_{ot}} \frac{1}{2} ||w||$$

s.t. 
$$Y(W_s^T X + W_{os}) + Y(W_t^T X + W_{ot}) > 1$$

To simply the equation, Let  $C_s = Y(W_s^T X + W_{os})$ . Then the constraint becomes:

$$Y(W_s^T X + W_{os}) \ge 1 - C_s$$

To get the optimized estimators, we use Lanrange Multiplier. That is:

$$L_p = \frac{1}{2}||w||^2 - \sum_{i=1}^N \alpha'[y_i(w^T x_i + w_o - (1 - C_s)]$$
(1)

Take derivate with respect to w and  $w_0$ , we will get

$$\frac{\partial L_p}{\partial w}: w - \sum \alpha_i y_i x_i = 0 \to w = \sum \alpha_i y_i x_i$$

$$\frac{\partial L_p}{\partial w_o} : 0 = \sum \alpha_i y_i$$

Substitute the partial derivative into (1), we will get the objective function to be:

$$\arg\min\frac{1}{2}\sum_{i}\sum_{j}\alpha_{i}\alpha_{j}y_{i}y_{j}x_{i}x_{j} + (1 - C_{s})\sum_{i}\alpha_{i}$$
(2)

Then we use the Quadratic Programming function (QP) to get the estimated  $\alpha$ . then obtain the resulting  $w_t$  and  $w_{ot}$ .

$$w_t = \sum \alpha_i y_i x_i$$
$$w_{ot} = -\frac{1}{2} (wx^+ + wx^-)$$

The above derivation is important for training the target data and use the fit function to predict the testing data. The added constraints will cause the QP function changed.

#### Our Algorithm:

- Step 1: Train the Source data and obtain a constant  $C_s = Y(W_s^T X + W_{os})$ . Where  $W_s$  and  $W_{os}$  are obtained from the QP optimization.
- Step 2: Train the Target data by adding  $C_s$  in the constraint part. The resulting Langrange Multiplier objective function has been derived above. Get the resulting  $w_t, w_{ot}$
- Step 3: Apply the estimated  $w_t, w_{ot}$  to the fit function  $f = sign(w_t x + w_{ot})$ . Use it along with the Testing data to get the predictions.

## Result: