Name: Jianyuan Lu GT ID: 903633385

# 1.SVM

(a)

Our optimization problem for SVM is:

|  |  |
| --- | --- |
|  |  |
|  | |

If we divide both sides of constrain function by , we can have:

We can denote the constrain function with new variables:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

So the original optimization problem becomes:

|  |  |
| --- | --- |
|  |  |
|  | |

(b)

The Lagrangian function is:

|  |  |
| --- | --- |
|  |  |

From the KKT condition, we can have: , so:

|  |  |
| --- | --- |
|  |  |
|  |  |

in the function is Lagrangian dual variables, which means

This expression for can be seen as the linear combination of input data times .

(c)

One of the KKT conditions is:

|  |  |
| --- | --- |
|  |  |

To satisfy this condition, there are two situations can occur:

|  |  |
| --- | --- |
|  |  |
|  |  |

From the geometry perspective, means . That is to say, the point doesn’t touch the dash line and is located away from the margin area. For those points, the weighted coefficient .

=0 means the point is located just at the dash line and in this case,. So those points will have the possibility to be considered when computing .

(d)

(i).

(ii).

No, the orientation will not change if the points is linearly separable. The orientation of margin decision boundary will always be parallel with the line linked (0,0) and (2,2).

# 2 Multi-class classification.

For KNN method, I set the number of nearest neighbors to be considered equals to 10. The confusion matrix, precision recall and F-1 score for testing are shown as below:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 978 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
|  |  | 13 | 1032 | 5 | 2 | 0 | 0 | 3 | 0 | 80 | 0 |
|  |  | 151 | 0 | 840 | 5 | 0 | 0 | 0 | 0 | 36 | 0 |
|  |  | 151 | 0 | 6 | 703 | 0 | 1 | 0 | 1 | 146 | 2 |
| Confusion matrix |  | 239 | 0 | 0 | 0 | 570 | 0 | 5 | 2 | 113 | 53 |
|  |  | 235 | 0 | 0 | 8 | 0 | 454 | 4 | 0 | 188 | 3 |
|  |  | 122 | 0 | 0 | 1 | 1 | 0 | 803 | 0 | 31 | 0 |
|  |  | 76 | 0 | 37 | 2 | 2 | 0 | 0 | 764 | 107 | 40 |
|  |  | 66 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 906 | 0 |
|  |  | 122 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 238 | 645 |

The precision, recall, F-1 score of KNN method are shown in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label | Precision | Recall | F1 score | Support |
| 0 | 0.45 | 1.00 | 0.62 | 980 |
| 1 | 1.00 | 0.91 | 0.95 | 1135 |
| 2 | 0.95 | 0.81 | 0.88 | 1032 |
| 3 | 0.97 | 0.70 | 0.81 | 1010 |
| 4 | 0.99 | 0.58 | 0.73 | 982 |
| 5 | 1.00 | 0.51 | 0.67 | 892 |
| 6 | 0.98 | 0.84 | 0.91 | 958 |
| 7 | 0.99 | 0.74 | 0.85 | 1028 |
| 8 | 0.49 | 0.93 | 0.64 | 974 |
| 9 | 0.87 | 0.64 | 0.74 | 1009 |

Besides, the total accuracy for KNN is 77%.

For logistic regression method, the confusion matrix, precision recall and F-1 score for testing are shown as below:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 968 | 0 | 0 | 2 | 0 | 0 | 4 | 1 | 5 | 0 |
|  |  | 0 | 1025 | 2 | 4 | 0 | 0 | 3 | 2 | 99 | 0 |
|  |  | 11 | 4 | 872 | 22 | 5 | 0 | 16 | 3 | 95 | 4 |
|  |  | 5 | 0 | 9 | 916 | 1 | 1 | 3 | 7 | 62 | 6 |
| Confusion matrix |  | 3 | 0 | 5 | 3 | 874 | 0 | 12 | 1 | 35 | 49 |
|  |  | 17 | 1 | 1 | 81 | 7 | 410 | 18 | 3 | 338 | 16 |
|  |  | 13 | 2 | 3 | 3 | 5 | 1 | 918 | 1 | 12 | 0 |
|  |  | 4 | 4 | 20 | 17 | 8 | 0 | 0 | 839 | 28 | 108 |
|  |  | 6 | 1 | 2 | 12 | 5 | 1 | 5 | 1 | 939 | 2 |
|  |  | 10 | 1 | 0 | 8 | 12 | 0 | 0 | 0 | 40 | 938 |

The precision, recall, F-1 score of KNN method are shown in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label | Precision | Recall | F1 score | Support |
| 0 | 0.93 | 0.99 | 0.96 | 980 |
| 1 | 0.99 | 0.90 | 0.94 | 1135 |
| 2 | 0.95 | 0.84 | 0.90 | 1032 |
| 3 | 0.86 | 0.92 | 0.88 | 1010 |
| 4 | 0.95 | 0.89 | 0.92 | 982 |
| 5 | 0.99 | 0.46 | 0.63 | 892 |
| 6 | 0.94 | 0.96 | 0.95 | 958 |
| 7 | 0.98 | 0.82 | 0.89 | 1028 |
| 8 | 0.57 | 0.96 | 0.71 | 974 |
| 9 | 0.84 | 0.93 | 0.88 | 1009 |

Besides, the total accuracy for neural network is 87%.

For linear SVM method, the confusion matrix, precision recall and F-1 score for testing are shown as below:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 968 | 0 | 1 | 2 | 1 | 1 | 4 | 0 | 2 | 1 |
|  |  | 0 | 1002 | 6 | 3 | 0 | 1 | 4 | 0 | 118 | 1 |
|  |  | 19 | 6 | 901 | 29 | 4 | 0 | 15 | 3 | 54 | 1 |
|  |  | 7 | 1 | 4 | 940 | 0 | 1 | 3 | 1 | 51 | 2 |
| Confusion matrix |  | 7 | 0 | 9 | 2 | 885 | 0 | 20 | 0 | 9 | 50 |
|  |  | 35 | 4 | 4 | 137 | 13 | 429 | 15 | 0 | 242 | 13 |
|  |  | 17 | 1 | 7 | 1 | 4 | 1 | 917 | 1 | 9 | 0 |
|  |  | 4 | 2 | 21 | 32 | 20 | 0 | 0 | 852 | 36 | 61 |
|  |  | 9 | 2 | 1 | 13 | 6 | 1 | 6 | 0 | 935 | 1 |
|  |  | 9 | 4 | 1 | 15 | 28 | 0 | 0 | 2 | 44 | 906 |

The precision, recall, F-1 score of KNN method are shown in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label | Precision | Recall | F1 score | Support |
| 0 | 0.90 | 0.99 | 0.94 | 980 |
| 1 | 0.98 | 0.88 | 0.93 | 1135 |
| 2 | 0.94 | 0.87 | 0.91 | 1032 |
| 3 | 0.80 | 0.93 | 0.86 | 1010 |
| 4 | 0.92 | 0.90 | 0.91 | 982 |
| 5 | 0.99 | 0.48 | 0.65 | 892 |
| 6 | 0.93 | 0.96 | 0.94 | 958 |
| 7 | 0.99 | 0.83 | 0.90 | 1028 |
| 8 | 0.62 | 0.96 | 0.76 | 974 |
| 9 | 0.87 | 0.90 | 0.89 | 1009 |

Besides, the total accuracy for neural network is 87%.

For neural network method, the confusion matrix, precision recall and F-1 score for testing are shown as below:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 953 | 0 | 3 | 2 | 2 | 10 | 6 | 2 | 2 | 0 |
|  |  | 0 | 1117 | 1 | 6 | 1 | 1 | 3 | 5 | 1 | 0 |
|  |  | 11 | 13 | 932 | 39 | 5 | 6 | 4 | 16 | 5 | 1 |
|  |  | 3 | 0 | 7 | 961 | 1 | 18 | 0 | 15 | 5 | 0 |
| Confusion matrix |  | 3 | 6 | 10 | 5 | 875 | 10 | 7 | 16 | 4 | 46 |
|  |  | 6 | 5 | 2 | 23 | 3 | 834 | 2 | 4 | 8 | 5 |
|  |  | 11 | 7 | 8 | 1 | 3 | 27 | 897 | 1 | 3 | 0 |
|  |  | 1 | 7 | 9 | 9 | 1 | 1 | 1 | 985 | 1 | 13 |
|  |  | 10 | 6 | 2 | 32 | 8 | 25 | 8 | 10 | 868 | 5 |
|  |  | 6 | 10 | 0 | 15 | 15 | 7 | 1 | 28 | 10 | 917 |

The precision, recall, F-1 score of KNN method are shown in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label | Precision | Recall | F1 score | Support |
| 0 | 0.95 | 0.97 | 0.96 | 980 |
| 1 | 0.95 | 0.98 | 0.97 | 1135 |
| 2 | 0.96 | 0.90 | 0.93 | 1032 |
| 3 | 0.88 | 0.95 | 0.91 | 1010 |
| 4 | 0.96 | 0.89 | 0.92 | 982 |
| 5 | 0.89 | 0.93 | 0.91 | 892 |
| 6 | 0.97 | 0.94 | 0.95 | 958 |
| 7 | 0.91 | 0.96 | 0.93 | 1028 |
| 8 | 0.96 | 0.89 | 0.92 | 974 |
| 9 | 0.93 | 0.91 | 0.92 | 1009 |

Besides, the total accuracy for neural network is 93%.

# 3 Neural networks.

(a).

For the logistic regression, we will evaluate the possibility for each label and assign the label as the output with the biggest possibility. The function is:

|  |  |
| --- | --- |
|  |  |
|  |  |

Usually if , we will assign label “1” as the output. If not, the output will become “0”.

As for neural network, if there is no hidden layer, only existing the input and output layer, the output will be computed as . If the value is larger than 0.5, the output is 1, otherwise output will be 0.

So these two methods will be the same in this circumstance.

(b).

The cost function is shown as follows:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Since is a two-dimensional vector, , . So the partial derivative with respect to are shown as follows:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Where is the corresponding element in with respect to .

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Where is the corresponding element in with respect to .