CSE4088 Introduction to Machine Learning

Homework 4 Report

SVM with Soft Margins

To solve problems in the homework, we apply soft-margin SVM to handwritten digits from the processed US Postal Service Zip Code data set. We will train two types of binary classifiers; one-versus-one (one digit is class +1 and another digit is class −1, with the rest of the digits disregarded), and one-versus-all (one digit is class +1 and the rest of the digits are class −1).

Implement SVM with soft margin on the above zip-code data set by solving

$$\min(\alpha) \quad \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m K(x_n, x_m) - \sum_{n=1}^{N}\alpha_n$$

$$s.t. \quad \sum_{n=1}^{N} y_n\alpha_n = 0, \qquad 0 \le \alpha_n \le C, \qquad n = 1, \dots, N$$

- **Polynomial Kernel**

Consider the polynomial kernel $K(x_n, x_m) = (1 + x_n^T x_m)^Q$, where ! is the degree of the polynomial.

2. With C = 0.01 and Q = 2, which of the following classifiers has the highest E_{in}?

Program Output

```
Question 2
C=0.010000, Q=2, classifier value=0, E_in=0.105884
C=0.010000, Q=2, classifier value=2, E_in=0.100261
C=0.010000, Q=2, classifier value=4, E_in=0.089425
C=0.010000, Q=2, classifier value=6, E_in=0.091071
C=0.010000, Q=2, classifier value=8, E_in=0.074338
```

According to program output, 0 versus all classifier value has highest E_{in} value.

The answer is [a].

3. With C = 0.01 and Q = 2, which of the following classifiers has the lowest E_{in}?

Program Output

```
Question 3
C=0.010000, Q=2, classifier value=1, E_in=0.014401
C=0.010000, Q=2, classifier value=3, E_in=0.090248
C=0.010000, Q=2, classifier value=5, E_in=0.076258
C=0.010000, Q=2, classifier value=7, E_in=0.088465
C=0.010000, Q=2, classifier value=9, E_in=0.088328
```

According to program output, 1 versus all classifier value has lowest E_{in} value.

The answer is [a].

4. Comparing the two selected classifiers from Problems 2 and 3, which of the following values is the closest to the difference between the number of support vectors of these two classifiers?

Program Output

```
Question 4
C=0.010000, Q=2, classifier value=0, E_in=0.105884
C=0.010000, Q=2, classifier value=1, E_in=0.014401
Difference between the number of support vectors of cls_0 and cls_1 = 1793
```

According to program output, the difference between support vectors counts is closed to 1800.

The answer is [c].

5. Consider the 1 versus 5 classifier with $Q = 2$ and $C \in \{0.001, 0.01, 0.1, 1\}$. Which of the following statements is correct? Going up or down means strictly so?

Program Output

```
Question 5
C          E_in      E_out      Support_vector_count
0.00100    0.00448    0.01651    76
0.01000    0.00448    0.01887    34
0.10000    0.00448    0.01887    24
1.00000    0.00320    0.01887    24
```

According to program output, minimum E_{in} is achieved with maximum C. Statements [a], [b] and [c] are incorrect, statement [d] is correct.

The answer is [d].

6. In the 1 versus 5 classifier, comparing $Q = 2$ with $Q = 5$, which of the following statements is correct?

Program Output

```
Question 6

C          E_in       E_out      Support_vector_count
0.00010    0.00897    0.01651    236
0.00100    0.00448    0.01651    76
0.01000    0.00448    0.01887    34
0.10000    0.00448    0.01887    24
1.00000    0.00320    0.01887    24

C          E_in       E_out      Support_vector_count
0.00010    0.00448    0.01887    26
0.00100    0.00448    0.02123    25
0.01000    0.00384    0.02123    23
0.10000    0.00320    0.01887    25
1.00000    0.00320    0.02123    21
```

First block of the program output shows the result of $Q = 2$. Second block of the program output shows the result of $Q = 5$. According to the results, when $C = 0.001$, the number of support vectors is lower at $Q = 5$ than at $Q = 2$.

The answer is [b].

- **Cross Validation**

In the next two problems, we will experiment with 10-fold cross validation for the polynomial kernel. Because E_{cv} is a random variable that depends on the random partition of the data, we will try 100 runs with different partitions and base our answer on how many runs lead to a particular choice.

7. Consider the 1 versus 5 classifier with $Q = 2$. We use $E_{cv}$to select $C \in \{0.0001, 0.001, 0.01, 0.1, 1\}$. If there is a tie in E_{cv}, select the smaller C. Within the 100 random runs, which of the following statements is correct?

Program Output

```
Question 7-8
Iteration 0, min support vector: 1.0000
Iteration 1, min support vector: 0.0010
Iteration 2, min support vector: 0.0010
Iteration 3, min support vector: 0.0100
...
Iteration 97, min support vector: 0.0010
Iteration 98, min support vector: 0.0100
Iteration 99, min support vector: 0.1000
C      Hits   Mean
0.0001 0     0.0098
0.0010 47    0.0048
0.0100 29    0.0047
0.1000 11    0.0048
1.0000 13    0.0049
Mean E_cv: 0.0058
```

When we count C values with minimum error, $C = 0.001$ has 47 times minimum error rate. The most often selected C value in 100 sample runs is $C = 0.001$.

The answer is [b].

8. Again, consider the 1 versus 5 classifier with $Q = 2$. For the winning selection in the previous problem, the average value of E_{cv} over the 100 runs is closest to which option?

According to result in question 7, average E_{cv} is 0.0058. It is closest to 0.005.

The answer is [c].

- **RBF Kernel**

Consider the radial basis function (RBF) kernel $K(x_n, x_m) = \exp(-||x_n - x_m||^2)$ in the soft-margin SVM approach Focus on the 1 versus 5 classifier.

9. Which of the following values of C results in the lowest E_{in}?

Program Output

```
Question 9-10
      C           E_in    E_out
      0.01        0.0038  0.0236
      1.00        0.0045  0.0212
    100.00        0.0032  0.0189
  10000.00        0.0026  0.0236
1000000.00        0.0006  0.0236
```

According to program output, lowest E_{in} is achieved when $C = 10^6$.

The answer is [e].

10. Which of the following values of C results in the lowest E_{out}?

According to the results in question 9, the lowest E_{out} is achieved when $C = 100$.

The answer is [c].