WORKSHOP 6 – Support Vector Machines

CMP9137M – MACHINE LEARNING

Overview

In this exercise, you are expected to understand the principle of support vector machines and use Python to train a SVM classifier.

Problem 1:

Assume you are given data points: $x_1 = 1, y_1 = +1; x_2 = 2, y_2 = +1; x_3 = 4, y_3 = -1, x_4 = 5, y_4 = -1; x_5 = 6, y_5 = +1$. Work on the following tasks:

1) Consider a hard margin SVM model; write down the original and dual optimization problems explicitly, i.e., inserting the data points into the optimization problems.
2) Consider a SVM model with the kernel function $K(x_i, x_j) = \left(1 + x_i^T x_j\right)^2$ in your hard margin SVM model. From a quadratic optimization solver, we have the solutions for the dual optimization problem: $\alpha_1 = 0, \alpha_2 = 2.5, \alpha_3 = 0, \alpha_4 = 7.333, \alpha_5 = 4.833$. For a new data point, the discriminant function (the classifier) is
   $\mathrm{f}(\mathrm{x}_*) = \sum_i \alpha_i y_i \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_*}) + b$

   a) Find b
   b) Write the discriminant function with those training data point explicitly
   c) Classify the new data points: $x_* = 4.5, 1.5$.

Solutions:

1) **For this particularly problem, SVM is to solve the following original optimization problem by inserting those data points into the max margin optimization problem:**

$$\boldsymbol{min\_w}\ ||\boldsymbol{w}||\text{^}\mathbf{2}$$
$$\boldsymbol{s.t.}, \boldsymbol{w} + \boldsymbol{w_0} \geq \mathbf{1}$$
$$\mathbf{2}\boldsymbol{w} + \boldsymbol{w_0} \geq \mathbf{1}$$
$$-\mathbf{1} \times (-\boldsymbol{w} + \boldsymbol{w_0}) \geq \mathbf{1}$$
$$-\mathbf{1} \times (\mathbf{5}\boldsymbol{w} + \boldsymbol{w_0}) \geq \mathbf{1}$$
$$\mathbf{6}\boldsymbol{w} + \boldsymbol{w_0} \geq \mathbf{1}$$

The dual formulation for this problem is:

$$\boldsymbol{max_\alpha} \sum_{i=1}^{5} \boldsymbol{\alpha_i} - \frac{\mathbf{1}}{\mathbf{2}} \times \mathbf{2} \times (\boldsymbol{\alpha_1\alpha_1} + \mathbf{4}\boldsymbol{\alpha_2\alpha_2} + \mathbf{16}\boldsymbol{\alpha_3\alpha_3} + \mathbf{25}\boldsymbol{\alpha_4\alpha_4} + \mathbf{36}\boldsymbol{\alpha_5\alpha_5} + \mathbf{2}\boldsymbol{\alpha_1\alpha_2}$$
$$- \mathbf{4}\boldsymbol{\alpha_1\alpha_3} - \mathbf{5}\boldsymbol{\alpha_1\alpha_4} + \mathbf{6}\boldsymbol{\alpha_1\alpha_5} - \mathbf{8}\boldsymbol{\alpha_2\alpha_3} - \mathbf{10}\boldsymbol{\alpha_2\alpha_4} + \mathbf{12}\boldsymbol{\alpha_2\alpha_5} + \mathbf{20}\boldsymbol{\alpha_3\alpha_4}$$
$$- \mathbf{24}\boldsymbol{\alpha_3\alpha_5} - \mathbf{30}\boldsymbol{\alpha_4\alpha_5})$$
$$\boldsymbol{s.t.}, \boldsymbol{\alpha_1} + \boldsymbol{\alpha_2} - \boldsymbol{\alpha_3} - \boldsymbol{\alpha_4} + \boldsymbol{\alpha_5} = \mathbf{0}, \boldsymbol{\alpha_i} \geq \mathbf{0}$$

2) **Insert all the training data points into the discriminant function:**

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_i \alpha_i y_i \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x}) + b$$
$$= \mathbf{2.5} \times \mathbf{1} \times \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_2}) + \mathbf{7.333} \times (-\mathbf{1}) \times \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_4}) + \mathbf{4.833} \times \mathbf{1} \times \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x_5}) + \boldsymbol{b}$$
$$= \mathbf{2.5} \times (\mathbf{2}\boldsymbol{x} + \mathbf{1})^2 - \mathbf{7.333} \times (\mathbf{5}\boldsymbol{x} + \mathbf{1})^2 + \mathbf{4.833} \times (\mathbf{6}\boldsymbol{x} + \mathbf{1})^2 + \boldsymbol{b}$$
$$= \mathbf{0.665}\boldsymbol{x}^2 - \mathbf{5.33}\boldsymbol{x} + \boldsymbol{b}$$

Since $\alpha_2, \alpha_4, \alpha_5$ are not zero, these training data points must lie on the margin. This means that for each of these data points, the following conditions are satisfied:
For x_2: $y_2(0.665x_2^2 - 5.33x_2 + b) = 1$; we obtain $b = 9$
For x_4: $y_4(0.665x_4^2 - 5.33x_4 + b) = 1$; we obtain $b = 9.03$
For x_5: $y_5(0.665x_5^2 - 5.33x_5 + b) = 1$; we obtain $b = 9.04$
Therefore $b = \frac{9+9.03+9.04}{3} = 9.02$.

We now can use the discriminant function to classify the new data points:
For $x_* = 4.5, f(4.5) = 0.665 * 4.5^2 - 5.33 * 4.5 + 9.02 = -1.499$.
Therefore, this data point is classified to negative class.
For $x_* = 1.5, f(1.5) = 0.665 * 1.5^2 - 5.33 * 1.5 + 9.02 = 2.521$.
Therefore, this data point is classified to positive class.

We can see that after using the polynomial kernel function, the 1-dimensional data (Figure 1) is mapped to 2-dimensional space (Figure 2), and in this example, the map defined by the kernel function is $K: x \rightarrow (x, x^2)$. We can see that in the original space, the data is not linearly separable, but after the mapping the data is linearly separable in feature space.

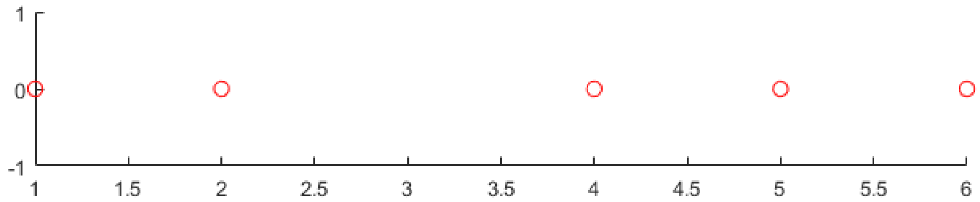

Figure 1: Training data in original space

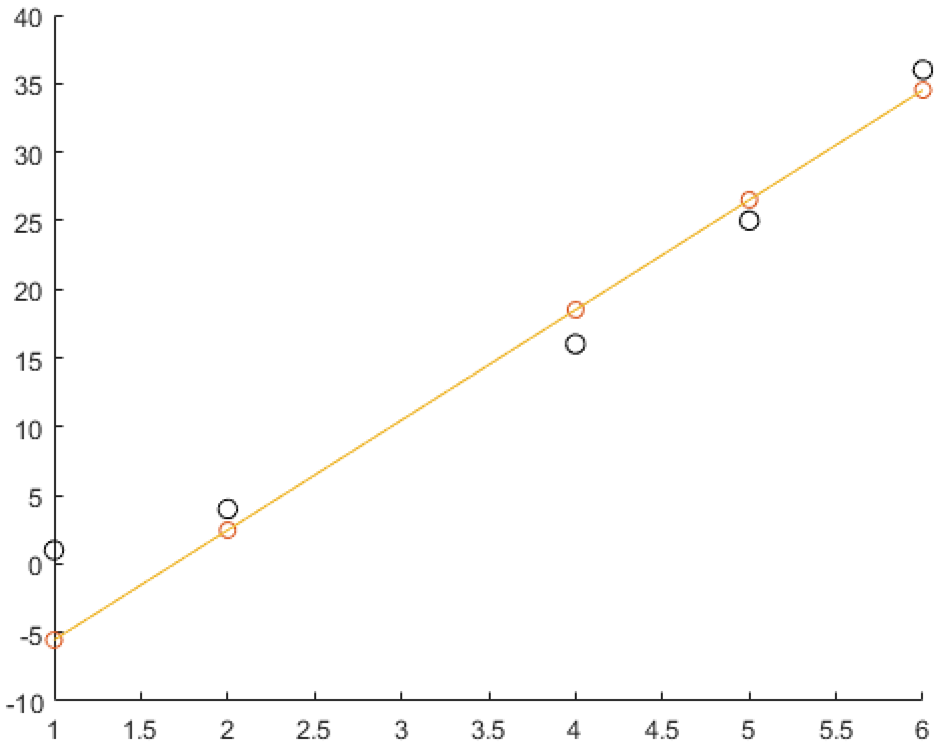

Figure 2: Training data in feature space; the yellow line is the decision boundary

Problem 2:

You are given a data set in a CSV file with the first two columns being features and the third column being the class labels. Work on the following tasks using Python or Matlab:

1) Randomly split the data to training (80%) and test (20%) sets.
2) Plot the scatter plots of the training data: positive examples in red and negative examples in blue circles.
3) Train a SVM classifier using the training data; you could use the “scikit-learn” module for Python or any other modules. You should plot the trained decision boundary. Can you point out which are the support vectors? What are the prediction results on the test data?

Solution: see the code demo_svm.py