THU-70250043-0, Pattern Recognition (Spring 2021) **Homework: 5**

SVM

Lecturer: *Changshui Zhang* zcs@mail.tsinghua.edu.cn

Hong Zhao vzhao@tsinghua.edu.cn

Student: xxx@mails.tsinghua.edu.cn

Problem 1

Given an unlabeled set of examples $\left\{x^{(1)}, ..., x^{(n)}\right\}$ the one-class SVM algorithm tries to find a direction ω that maximally separates the data from the origin.(This turns out to be useful for anomaly detection.)

More precisely, it solves the (primal) optimization problem:

$$\begin{aligned} \min_{\omega} \quad & \frac{1}{2}\omega^T\omega \\ \text{s.t.} \quad & \omega^T x^{(i)} \geq 1, \qquad i = 1, \dots, n \end{aligned}$$

A new test example x is labeled 1 if $\omega^T x \geq 1$, and 0 otherwise

1.1 The primal optimization problem for the one-class SVM was given above. Write down the corresponding dual optimization problem. Simplify your answer as much as possible. In particular, ω should not appear in your answer.

1.2 Can the one-class SVM be kernelized (both in training and testing)? Justify your answer.

Problem 2

Consider a dataset with 2 points in 1d: $x_1 = 0$, $x_2 = \sqrt{2}$ with labels $y_1 = -1$, $y_2 = 1$. Consider mapping each point to 3d using the feature vector $\Phi = [1, \sqrt{2}x, x^2]^T$. (This is equivalent to using a second order polynomial kernel.) The max margin classifier has the form:

$$\begin{aligned} min||w||^2 \quad & s.t. & (1) \\ y_1(w^T\Phi(x_1) + b) \quad & \geq \quad 1 & (2) \\ y_2(w^T\Phi(x_2) + b) \quad & \geq \quad 1 & (3) \end{aligned}$$

a. Write down a vector that is parallel to the optimal vector w.

b. What is the value of the margin that is achieved by this w? Hint: recall that the margin is the distance from each support vector to the decision boundary. Hint 2: think about the geometry of 2 points in space, with a line separating one from the other.

c. Solve for w, using the fact the margin is equal to $1/||w||$.

d. Write down the form of the discriminant function $f(x) = w^T\Phi(x) + b$ as an explicit function of x.

Problem 3

The exclusive-OR is the simplest problem that cannot be solved using a linear discriminant operating directly on the features. The points $k = 1, 3$ at $\mathbf{x} = (\mathbf{1}, \mathbf{1})^\mathbf{t}$ and $\mathbf{x} = (-\mathbf{1}, -\mathbf{1})^\mathbf{t}$ are in category ω_1 (red in Figure 1), while $k = 2, 4$ at $\mathbf{x} = (\mathbf{1}, -\mathbf{1})^\mathbf{t}$ and $\mathbf{x} = (-\mathbf{1}, \mathbf{1})^\mathbf{t}$ are in ω_2 (black in Figure 1). Following the approach of SVM, we use kernel function to map the features to a higher dimension space where they can be linearly separated.

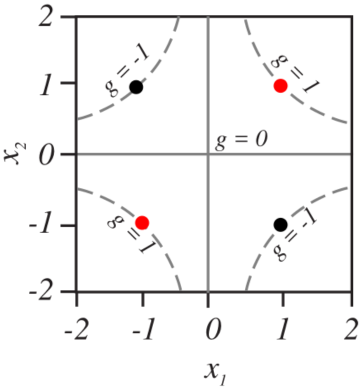

Figure 1: The XOR problem in the original (x_1, x_2) feature space.

3.1 Consider the polynomial kernel of degree 2:

$$K(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i}^\mathbf{t}\mathbf{x_j} + \mathbf{1})^\mathbf{2},$$

where $\boldsymbol{x_i} = (x_{i1}, x_{i2})^t$ and $\boldsymbol{x_j} = (x_{j1}, x_{j2})^t$.

Show that it corresponds to mapping

$$\Phi(x_1, x_2) = [1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2].$$

3.2 Derive the dual problem in the 6-dimensional space with Lagrange multipliers $\alpha_i, i = 1, 2, 3, 4$ as the only variants.

3.3 Solve the dual problem analytically (without programming). What are the support vectors? *Hint: use the symmetry of the problem.*

3.4 Derive the final discriminant function $g(\mathbf{x})$ and the decision hyperplane.

3.5 Plot the data points on the subspace of $(\sqrt{2}x_1, \sqrt{2}x_1x_2)$. Demonstrate the decision hyperplane and the margin in your plot.

Problem 4

Write a program to complete the following task adopting the SVM algorithm (you could use some toolkits or source code). Train a SVM classifier with data from ω_1 and ω_2 in the following table. Preprocess each training pattern to form a new vector having components $1, x_1, x_2, x_1^2, x_1x_2$ and x_2^2.

Category	x_1	x_2
ω_1	-3.0	-2.9
ω_1	0.5	8.7
ω_1	2.9	2.1
ω_1	-0.1	5.2
ω_1	-4.0	2.2
ω_1	-1.3	3.7
ω_1	-3.4	6.2
ω_1	-4.1	3.4
ω_1	-5.1	1.6
ω_1	1.9	5.1
ω_2	-2.0	-8.4
ω_2	-8.9	0.2
ω_2	-4.2	-7.7
ω_2	-8.5	-3.2
ω_2	-6.7	-4.0
ω_2	-0.5	-9.2
ω_2	-5.3	-6.7
ω_2	-8.7	-6.4
ω_2	-7.1	-9.7
ω_2	-8.0	-6.3

4.1 Train you classifier with just the first point in the ω_1 and ω_2 and find the separating hyperplane and the margin.

4.2 Repeat 4.1 with the first two points in each category (four points total), the first three point and so on, until the transformed patterns cannot be linearly separated in the transformed space.

Hint: You needn't program the SVM algorithm by yourself, you can just use some toolkits or source code such as libsvm `http://www.csie.ntu.edu.tw/~cjlin/libsvm/` for MATLAB or scikit-learn `https://scikit-learn.org/stable/modules/svm.html#svm-classification` for python. You should declare the toolkit you used in your project.