SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.007 Machine Learning, Summer 2022

Homework 3: Classification with Support Vector Machines

Due: 1st of July 2022
This homework will be graded by TA.

1 Support Vector Machines

Question 1.1 [20 pts]
Given the mapping $x = [x_1 \;\; x_2]^T \rightarrow \varphi(x) = [1 \;\; x_1^2 \;\; \sqrt{2}x_1x_2 \;\; x_2^2 \;\; \sqrt{2}x_1 \;\; \sqrt{2}x_2]^T$
a) Determine the Kernel $K(x, y)$ [*10pts*]
b) Calculate the value of the Kernel $K(x, y)$ if $x = [1 \;\; 2]^T$ and $y = [3 \;\; 4]^T$ [*10pts*]

Question 1.2 [30 pts]
The primal problem of SVM with soft margin is given below:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i \\ \text{subject to} \quad & d_i(w^T x_i + b) - 1 + \xi_i \geq 0\,, \quad \xi_i \geq 0 \end{aligned}$$

1) Using Lagrange multipliers and KKT conditions, can you derive the formulation of dual problem with soft margin? Please note that the dual form is already provided in slides, so we expect you to go through the mathematical steps. [*20pts*]
2) Explain in which cases we would prefer to use soft margin rather than hard margin. [*10pts*]

Question 1.3: Hands-on [50 pts]
Download and install the widely used SVM implementation LIBSVM `https://github.com/cjlin1/libsvm`, or `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`. We expect you to install the package on your own – this is part of learning how to use off-the-shelf machine learning software. Read the documentation to understand how to use it.

Download fishorrock folder. In that folder there are training.txt and test.txt, which respectively contain 145 training examples and 63 test examples in LIBSVM format. Sonar systems are generally used underwater for range finding, classification and detection. In the fishing industry, a sonar is generally used to detect (or classify) fish, and the seafloor around the vessel. The task in this homework is to train an SVM model to discriminate between signals, and to classify if the object is a fish or rock given 60 attributes about the signal. Please note that this is a binary classification task.

Run LIBSVM to classify objects with the following kernels:

- linear
- polynomial
- radial basis function
- sigmoid

You should use default values for all other parameters. Please report the test accuracy for each kernel. Which kernel would you choose and why?