CS 229 Machine Learning, spring 2019

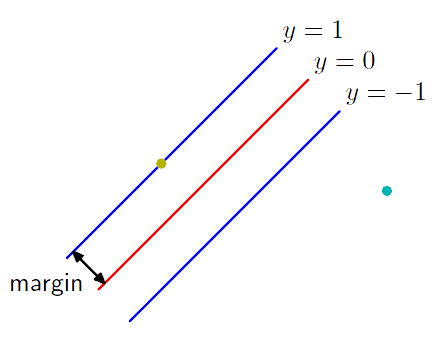
Homework 7:

Support Vector Machine

Due Saturday May 4, 11:59pm

Submit by the **blackboard system**

**Question1: (30 points) *Margin* for the maximum-margin hyper-plane (Exercise 7.4 of Bishop’s book, 0 pt for solutions copied from internet)**



ᵨ

Show that the value ρ of the margin for the maximum-margin hyperplane is given by

where are given by maximizing

subject to

and

**Question2: (70 points) Classification by SVM**

**Code:**

You can use any SVM tools.

**Data:**

You can download the data from

<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

Take either the red-wine or the white-wine data set.

Take “quality” as class label, e.g., 1-5 as **negative**, while 6-10 as **positive**.

Or you can use some of your own challenging dataset for classification.

**Evaluation and Testing:**

Divide the whole data set into training data and testing data, e.g., 60% for training and 40% for testing.

Use 5-fold Cross Validation for setting parameters, e.g., C and kernel parameters.

**Questions to answer:**

We have 3 different types of models for learning classifiers (SVM with 3 different types of kernel):

1 SVM with linear kernel

2 SVM with polynomial kernel

3 SVM with the radial basis function kernel

1. (15 points) For each learning model (each type of kernel),

use 60% of data for training SVM model (with the default parameters), and use the remaining 40% for testing.

1. Report the number of support vectors
2. Plot the ROC curve of testing results by ranking the decision values (3 curves in one figure)
3. Compute the AUC (Area Under Curve), which kernel is the best?

1. (45 points) For each learning model (each type of kernel),

Use 5-fold cross validation for setting the parameters of training process. Please note different kernels may have different parameters to set.

After cross validation, choose the best parameter setting, train the model by 60% of data again (the same data used in (1)), test the model by the remaining 40% of data.

1. Report the setting of parameters
2. Report the number of support vectors
3. Plot the ROC curve of testing results by ranking the decision values (3 curves in one figure)
4. Compute the AUC (Area Under Curve), which kernel is the best?
5. (10 points) make a table for comparing the AUC of different kernels with different setting of parameters (totally 6 AUC values), report the 3 best models (decided by their AUC values).