Problem Set 2

CS 4375

Due: 10/06/2025 by 11:59pm

Note: all answers should be accompanied by explanations and code for full credit. Late homeworks will not be accepted.

Problem 1: Feature Maps and Support Vector Machines (30 pts)

For this problem, consider the data set (mystery.data) attached to this homework. This data set contains four numeric attributes per row and the fifth entry is the class variable (either +1 or -1).

In this problem your goal is to find a perfect classifier for this data set that will generalize well to unseen data observations using support vector machines a feature maps. The quality of your solution will be determined by applying your method to unseen data. What to turn in:

1. Design a feature map than will make this data linearly separable and solve the *primal* SVM problem to find the max margin classifier in your feature space. You should report your learned parameters (the weights and bias), the optimal margin, and the support vectors.
2. A function `eval` in Python or MATLAB that takes as input a dataset (treated as a two-dimensional array) in the above form excluding the last column (i.e., a matrix with four numeric attributes per row). The function can apply whatever feature transformation you would like, and using whatever weights you learned, it should return a class label for each row of the input matrix as a vector.

A total of ten points for this problem will be allocated to the performance of your `eval` function on the unknown test set. These 10 points will be awarded by multiplying your accuracy on the unseen data times 10 and rounding up, i.e., if you get 79% of the labels correct, you will receive 8 points. If the TA is unable to run your code, you will receive zero points. No external ML libraries for SVMs can be used in your solution (you should use an external library for the quadratic programming solver, e.g., cvxopt for Python).

Problem 2: Modified SVM with Slack (30 pts)

Recall, in class, that we saw the SVM with slack formulation in which violation of the constraints was penalized linearly, e.g., $c\sum_i \xi_i$ for some hyperparameter $c \geq 0$. If instead, violation of the constraints is penalized quadratically, e.g., $c\sum_i \xi_i^2$, the resulting optimization problem is still convex.

1. What is the dual of this new slack formulation?
2. Explain how to construct a solution to the primal problem given a solution to the dual problem.

3. Can the kernel trick still be used?

4. Explain how to solve the primal problem with (sub)gradient descent.

Problem 3: SVMs on Real Data (40 pts)

For this problem, you will use the `magic.data` data set provided with this problem set. These data sets were generated using the Magic data set. Note that the class label is the last column in the data set (in this case zero or one). In the following questions, you should use the first 1000 rows for training, the next 1000 for validation, and the remaining rows for testing.

Again, no external ML libraries for SVMs can be used in your solution (you should use an external library for the quadratic programming solver, e.g., cvxopt for Python).

1. Primal SVMs

   (a) Use the SVM with slack formulation to train a classifier for each choice of $c \in \{10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$.

   (b) What is the accuracy of the learned classifier on the training set for each value of c?

   (c) What is the accuracy on the validation set for each value of c?

   (d) Train a new model by combining the training and validation data using the best hyper-parameter value, and report your accuracy on the test set.

2. Dual SVMs with Gaussian Kernels

   (a) Use the dual of the SVM with slack formulation to train a classifier for each choice of $c \in \{10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$ using a Gaussian kernel with $\sigma \in \{.01, .1, 1, 10, 100, 1000\}$.

   (b) What is the accuracy of the learned classifier on the training set for each pair of c and σ?

   (c) What is the accuracy on the validation set for each pair of c and σ?

   (d) Train a new model by combining the training and validation data using the best hyper-parameter values, and report your accuracy on the test set.

3. Modified SVMs with Slack

   (a) Use the modified SVM with slack formulation from Problem 2 each choice of $c \in \{10^{-1}, 10^0, 10^1, 10^2, 10^3, 10^4, 10^5\}$.

   (b) What is the accuracy of the learned classifier on the training set for each c?

   (c) What is the accuracy on the validation set for each c?

   (d) Train a new model by combining the training and validation data using the best hyper-parameter value, and report your accuracy on the test set.

4. Which of these approaches (if any) should be preferred for the classification task? Explain.