EE 518: Machine Learning Theory & Algorithms	Spring 2024

Homework 6
Due: May 31, 2024, 11:59PM PT

Student Name: *Instructor Name: John Lipor*

Problem 1   Kernel SVM (15 pts)

Your task in this problem is to implement the `SGD for Solving Soft-SVM with Kernels` algorithm given on pg. 223 of UML and test this algorithm on `prob1.py`. While the algorithm appears to require actually mapping the feature vectors to a high-dimensional space, this is actually not necessary for prediction. Note that prediction is performed via the equation

$$\hat{y}_i = \text{sign}\left(\bar{w}^T \psi\left(x_i\right)\right) = \text{sign}\left(\sum_{j=1}^{m} \bar{\alpha}_j \psi\left(x_j\right)^T \psi\left(x_i\right)\right),$$

and hence the predictions can also be kernelized. Use the Gaussian kernel with your choice of values for σ and λ, and use $T = 10,000$ iterations. **Do not** use any loops when computing your kernel matrices or your predictions. **Turn in** your code, your selected values of σ and λ, as well as the training and test errors. *Note:* Your final test error should be below 5%.

Problem 2   DSS: Regression Models (10 pts, 5 pts, 5 pts)

(*DSS rules apply.*) Our focus in this class has been largely on classification problems. However, as you saw with (not-so-mini) Mini Project 2, the models we have studied frequently have variants that can be used for regression as well. In this problem, you will utilize three of the models we have studied to perform regression on the California housing dataset.

(a) Use the `sklearn` support vector regression implementation to predict on the housing dataset as loaded in `prob2.py`. There are several kernels available, as well as a number of parameters to select for each kernel. You should use model selection to decide which kernel and parameters are most appropriate. Note that you may choose to either use a validation set or cross validation, but you should justify your choice. Further, the performance of SVR will be greatly improved by performing data preprocessing, for which you can make use of the `sklearn.preprocessing` library. You should also decide on the appropriate preprocessing using *only the training data*, then apply your chosen preprocessing to the entire dataset (training and test) simultaneously. **Turn in** a description of your approach to model selection, your approach to preprocessing, your selected SVR parameters, your training and test errors, and your code.

(b) Compare your SVR errors to ridge regression and k-nearest neighbors (you may use the `sklearn` implementation of these). Use the same data preprocessing for all three algorithms, but be sure to tune parameters for each algorithm individually. **Turn in** your training and test errors for ridge regression and k-nearest neighbors.

(c) Provide a brief (half page maximum) description of how SVR works. Assume your audience is familiar with SVMs for classification and cite any sources you used to gain your understanding.

Problem 3   Positive Definite Kernels (10 pts)

Lemma 16.2 states that a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a valid kernel function if and only if the resulting Gram matrix is positive semidefinite. In the proof the authors claim that it is "trivial to see that if K implements an inner product in some Hilbert space then the Gram matrix is positive semidefinite." As we all know, one textbook author's trivial is often a student's homework problem. Prove this "trivial" claim.

Problem 4   SLT (5 pts)

(*SLT rules apply.*) UML, Ch. 15, Exercise 1. State how long you worked on the problem before looking at the solution.