2016 Fall CPS 571/STA 561: Homework 4

Duke University

October 23, 2016

1 Logistic Regression and Kernels

Consider l_2 regularized logistic regression with loss function

$$\mathcal{L}(f, \theta_0) = \sum_{i=1}^{n} \ln(1 + \exp(-y_i(f(\mathbf{x}_i) + \theta_0)) + \lambda \|f\|_{\mathcal{H}}^2.$$

where $y_i \in \{-1, 1\}$, $f(\mathbf{x}) = \boldsymbol{\theta}^\top \mathbf{x}$, and $\|f\|_{\mathcal{H}}^2 = \boldsymbol{\theta}^\top \boldsymbol{\theta}$.

(a) Define the reproducing kernel Hilbert space $\mathcal{H}$. Using the representer theorem, what do you know about the optimal solution?

(b) Given logistic loss function $g(\zeta) = \ln(1 + \exp(-\zeta))$, plot $\frac{1}{\ln 2} g(\zeta)$ and hinge loss function to compare them.

(c) We can rewrite the l_2 regularized logisitic regression as the following primal problem

$$\min_{\boldsymbol{\theta}, \theta_0, \boldsymbol{\zeta}} \frac{1}{2} \|\boldsymbol{\theta}\|^2 + C \sum_{i=1}^{n} g(\zeta_i)$$

subject to

$$y_i(f(\mathbf{x}_i) + \theta_0) \geq \zeta_i, \forall i$$

Derive the dual formulation by KKT (you are not required to solve it). Compare your result with non-separable SVM, and explain the similarities and differences.

2 SVM – Properties of the Maximum Margin Hyperplane

(a) Show analytically that a data set consisting of just two data points, one from each class, is sufficient to determine the location of the maximum-margin hyperplane. (You should only consider the separable case.)

(b) Argue that finding the maximum margin hyperplane is a convex optimization problem.

3 SVM Experiments

In this problem, you will experiment with the support vector machine (SVM), and various kernel functions.

(a) Implement a "hard" maximum-margin SVM classifier in MATLAB, R, or Python. Here, "hard" means that if the data are separable the SVM will return a maximum margin solution, but if the data are not separable, the code will fail. Your implementation should optimize the dual problem using a quadratic program solver or a specialized solver, and include a *train* function and a *predict* function. (Note that you do not need to write the solver yourself!)

(b) Train an SVM classifier with the kernel function $k(\mathbf{x}, \mathbf{z}) = \mathbf{x}^\top \mathbf{z}$ on 9/10ths of the credit card data set. (Randomly choose which tenth to leave for testing.) What is the accuracy of this classifier on the test data set? Show the ROC curves, and also report the AUC.

(c) Train an SVM classifier with the radial basis kernel

$$k(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|_2^2}{\sigma^2}\right)$$

on the credit card data training set, for $\sigma^2 = 2$ and $\sigma^2 = 20$. What is the accuracy of these classifiers on the test data set? Show the ROC curves, and also report the AUC.