STA208: Homework 2

Prof. James Sharpnack

Due 4/11 in class

In the following, show all your work. Feel free to do all the analytical questions first and then include the code and output second, but the different parts and which question that you are answering should be clearly marked. Code should be as modular as possible, points will be deducted for code that is not reusable (i.e. not broken into general purpose functions), and in the case of gratuitous hard coding.

1. The following losses are used as surrogate losses for large margin classification. Demonstrate if they are convex or not, and follow the instructions.
   (a) exponential loss: $\phi(x) = e^{-x}$
   (b) truncated quadratic loss: $\phi(x) = (\max\{1 - x, 0\})^2$
   (c) hinge loss: $\phi(x) = \max\{1 - x, 0\}$
   (d) sigmoid loss: $\phi(x) = 1 - \tanh(\kappa x)$, for fixed $\kappa > 0$
   (e) Plot these as a function of x.

   (This problem is due to notes of Larry Wasserman.)

2. Consider the least-squares problem with n p-dimensional covariates, $\{\mathbf{x}_i, y_i\}_{i=1}^n \subset \mathbb{R}^p \times \mathbb{R}$. We would like to fit the following linear model, $\hat{y}(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}$. Also, suppose that there are coefficients $C_+ \subset \{1, \ldots, p\}$ such that for all $j \in C_+$ we require that $\beta_j \geq 0, j \in C_+$, and another set $C_- \subset \{1, \ldots, p\}$, such that $\beta_j \leq 0, j \in C_-$ (assume that C_+ and C_- are non-overlapping). Suppose that $\mathbf{X}^\top \mathbf{X}$ is invertible.

   Such examples occur in insurance applications: the cost of a given insurance policy is based on a model for the amount of money a customer will cost the company, and each covariate is a variable specific to the customer (such as gender, age, credit history, etc.). It looks bad for the company if the insurance policy is more expensive for a customer that has an older account with the company than a newer account, when everything else is held fixed. Let $x_{i,j} = 1\{\text{customer } i \text{ is has had a policy for more than 2 years}\}$, then $\beta_j \geq 0$ is necessary for this property to hold.
   (a) Write the constrained optimization for the empirical risk minimization with the constraints.
   (b) Derive the dual for the optimization as a function of dual parameters.
   (c) Write the KKT conditions and remark on the implication of the complementary slackness condition.

3. Look at the dataset which can be found here: `https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center`
   (a) You will predict the final column in the dataset, which is an indicator if the person has made a blood donation. Form three different kernel functions and $n \times n$ kernel matrices of $K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$ that you think might be appropriate. You may use 2 different built in kernels, but must define one yourself.
   (b) Apply kernel SVMs and k-nearest neighbors (where the distance is $d(i,j) = k(\mathbf{x}_i, \mathbf{x}_i) + k(\mathbf{x}_j, \mathbf{x}_j) - 2k(\mathbf{x}_i, \mathbf{x}_j)$) with each of the 3 kernels.
   (c) Tune any parameters based on what you have learned about validation, and compare these methods with test errors (there are 6 different methods to compare, kNN and SVM with each kernel).