STA208: Homework 3

Prof. James Sharpnack

Due 4/25 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. This question is about your final project.
   (a) List the members of your group.
   (b) What is the title of your project?
   (c) Give a brief description of what you intend to do.
2. Consider the following linear classification loss function:
   $$\ell(\boldsymbol{\beta}; y_i, \mathbf{x}_i) = (1 - y_i \mathbf{x}_i^\top \boldsymbol{\beta})_+^2$$
   where $z_+ = \max\{0, z\}$. Consider the modified support vector machine program,
   $$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \ell(\boldsymbol{\beta}; y_i, \mathbf{x}_i) + \lambda \|\boldsymbol{\beta}\|_2^2. \tag{1}$$
   (a) Compute the gradient of the objective in (1) and verify that it is continuous.
   (b) Write the pseudocode for gradient descent with exact line search for this gradient (you do not need to write out the bisection algorithm for exact line search, only state what exact line search is.).
   (c) Suppose that we make a transformation $\mathbf{z}_i = \Phi(\mathbf{x}_i)$ and we use this as the design matrix. Rewrite the minimization for (1) with the generalized kernel trick (Hint: make the substitution $\boldsymbol{\beta} = \mathbf{Z}^\top \boldsymbol{\gamma}$ and write it as a function of $\boldsymbol{\gamma}$ and define $\mathbf{K} = \mathbf{Z}\mathbf{Z}^\top$.).
3. Consider the following way in which we encode logical statements. For $\mathbf{x} \in \{0,1\}^p$, we will encode the "and" and "or" statements: ($x_i > 0$ and $x_j > 0$) is the same as ($x_i x_j > 0$); ($x_i > 0$ or $x_j > 0$) is the same as ($x_i + x_j > 0$).
   (a) Suppose that $y_i = 1$ if and only if ($x_1 = 1$ and $x_2 = 1$) or $x_3 = 1$. Write a function $f(\mathbf{x})$ such that $f(\mathbf{x}) = 1$ if and only if y_i.
   (b) Define an embedding of $\mathbf{x}$ such that the dataset has a separating hyperplane, and write an equation for the separating hyperplane as a function of $\mathbf{z}_i = \Phi(\mathbf{x}_i)$.
   (c) From this embedding construct the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$.
   (d) Suppose that $y_i = 1$ if and only if ($x_1 = 1$ or $x_2 = 1$) and ($x_3 = 1$ and $x_4 = 1$). Write a function $f(\mathbf{x})$ such that $f(\mathbf{x}) > 0$ if and only if y_i.
   (e) Define an embedding of $\mathbf{x}$ such that this new dataset has a separating hyperplane, and write an equation for the separating hyperplane as a function of $\mathbf{z}_i = \Phi(\mathbf{x}_i)$.
   (f) From this embedding construct the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$.
   (g) Define the degree of a logical statement as the maximum number of variables involved in any monomial after we apply the substitutions defined above. Give an embedding and kernel that separates a dataset where $y_i = 1$ if and only if some logical statements of degree 3 is true.

4. Download the training and test datasets from the website for the Reuters data. The first column is the response variable and the rest are the counts of each term in the dictionary for each document.
   (a) Tune a linear SVM with the raw counts as the design matrix and tune with 5-fold cross validation.
   (b) Apply the tf-idf transformation and then tune a linear SVM with this as the design matrix, again with 5-fold cross validation.
   (c) Tune logistic regression to these and tune it with 5-fold cross validation.
   (d) Construct scores for each of these methods for the test set (this should be $\mathbf{x}_i^\top \boldsymbol{\beta}$), and construct the ROC and PR curves for them.