2016 Fall CPS 571/STA 561: Homework 3

Duke University

October 1, 2016

1 Convex Hull

For a set of data points $\{\mathbf{x}_n\}$, the convex hull of $\{\mathbf{x}_n\}$ is defined to be the set of all points which can be expressed as a linear combination of the points in $\{\mathbf{x}_n\}$, provided that the coefficients of the combinations add up to 1. Formally, the convex hull is the set of points given by

$$\mathbf{x} = \sum_n a_n \mathbf{x_n}$$

where $a_n \geq 0$ and $\sum_n a_n = 1$.

Recall that two sets of points $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_m\}$ are linearly separable if there exist any $\mathbf{w}$ and scalar offset w_0 such that $\mathbf{w}^\top \mathbf{x}_n + w_0 > 0$ and $\mathbf{w}^\top \mathbf{y}_m + w_0 < 0$. Show that if the convex hulls of any two sets of data points $\{\mathbf{x}_n\}$ and $\{\mathbf{y}_n\}$ intersect, then the two sets of points cannot be linearly separable.

2 "Cost sensitive" Adaboost

Consider a "cost sensitive" version of AdaBoost where some examples are more important than others. In particular, observations 1 through 10 are each worth twice what each of the other examples is worth. Thus, we decide to minimize a weighted objective function

$$\sum_{i=1}^{10} 2e^{-(\mathbf{M}\boldsymbol{\lambda})_i} + \sum_{i>10} e^{-(\mathbf{M}\boldsymbol{\lambda})_i}$$

This changes the AdaBoost algorithm. Write the new version of AdaBoost.

3 Margin Matrix of AdaBoost

Assume the weak learning assumption holds. Consider the matrix of margins $\mathbf{M} \in \{-1, 1\}^{m \times n}$ where $M_{ij} = y_i h_j(\mathbf{x}_i)$, and m is the size of dataset and n is the size of classifier family $\mathcal{H}$. Is there necessarily a linear classifier (meaning a linear combination of the weak classifiers) that classifies this dataset perfectly? Will AdaBoost find it?

Consider a symmetric dataset, where for each weak classifier, it's "twin" is included. That is, if $h_(x)$ is a weak classifier, then $-h(x)$ is also a weak classifier. If the weak learning assumption does not hold, can AdaBoost necessarily find a classifier that classifies the symmetric dataset perfectly? If so, explain why this is always true. If not, provide a symmetric matrix of margins corresponding to a dataset where there is no separating hyperplane.

4 Empirical Comparison

Perform an empirical comparison between at least 5 different machine learning algorithms for classification on the dataset of your choice from the UCI machine learning repository (http://archive.ics.uci.edu/ml/). You may use the skeleton code we provided. You can run algorithms that we have not covered in class yet. (You do not need to understand how these algorithms work in order to run someone else's code for these algorithms.) Perform 10 fold nested cross validation for one parameter for at least 2 different algorithms. Perform cross validation to evaluate the results. Please provide:

- Code. Make sure it is uploaded properly. Some students had problems with uploading code for the last assignment, so double check.
- Any pre-processing that you did.
- The list of algorithms you chose, the parameters you chose to optimize, and the values for those parameters you tried.
- ROC Curves for each individual feature, on all 10 of the cross validation "test" folds.
- ROC Curves for each algorithm for each of the 10 test folds. Create the ROC Curves using the first method from class, where you use only a single model to create the whole ROC curve.
- AUC values: report mean $\pm$ standard deviation over the 10 folds.

- A brief discussion: (i) Did the algorithms all perform similarly or did you see performance differences? (ii) Why do you think you obtained the results you got? (iii) Was the result of combining features better than the individual features alone?

Note that if the data set you choose is too big, you are permitted to choose a random subset of the observations. If you choose a dataset with categorical features, you can use "dummy variables," which means you would transform the categories into 0/1 features.