COMP 790-134 Machine Learning with Discriminative Methods—Homework 1

January 13, 2015

The objective of this assignment is to become more familiar with Matlab as a tool for exploring machine learning and to build some intuition for the underlying optimization problems involved in training a discriminative classifier. Due Thursday, January 15 before class.

1 Part 1 (Submit this part on paper.)

- Construct a set of around 10 labeled points in one dimension, ordered pairs (x_i, y_i) with $x_i \in \mathbb{R}^1$ and $y_i \in \{-1, 1\}$, and plot the $0-1$ loss as a function of the decision boundary x.
- Now construct a set of labeled points (still in 1D) so that the $0-1$ loss has multiple, disconnected, local minima. Show the points and plot the loss as a function of x.
- Write an algorithm to find the decision point with the minimal loss. What is the complexity of this algorithm in terms of the number of examples? Do other factors matter for complexity?
- How does this change for the hinge-loss?

2 Part 2 (Submit link to a web-page with your code and result images.)

- Write code to generate random binary (-1,1) labeled 2D data in a fixed range with an offset between the mean of the positive and negative data points. Plot these points.
- Write code to compute the $0-1$ loss of decision boundaries defined by a 2D vector w and boas term b so that the output of the decision function is $\text{sign}(w_1x_1+w_2x_2+b)$. Plot the loss as a function of w_1 and w_2. (For each value of w_1 and w_2 you should find the b that results in the smallest loss.) Plot the result as a heat map as shown in class. For fun, try looking at the heat map using something like `surf(lossmap); axis vis3d; rotate3d on;`
- Repeat the above using the hinge loss.
- How are the results with $0-1$ loss and hinge loss different?

3 Notes

To make sure the loss functions are clear. If you have data items (x_i, y_i) with $x_i \in \mathbb{R}^n$ a vector in n dimensions and $y_i \in \{-1, 1\}$ the corresponding binary label, and a function $f : \mathbb{R}^n \mapsto \mathbb{R}$, then on m data items $\{(x_i, y_i)\}_{i=1...m}$,

- the 0-1 loss is $\sum_{i=1}^{m} \frac{1}{2}(1 - \text{sign}(f(x_i))y_i)$ and
- the hinge loss is $\sum_{i=1}^{m} \max(0, 1 - f(x_i)y_i)$.