ML CS726 Fall'11 Project 2 Complex Classifiers

Angjoo Kanazawa, Ran Liu, Austin Myers

October 23, 2011

1 Gradient Descent and Linear Classification

1.1 WU1

Find a few values of step size where it converges and a few values where it diverges. Where does the threshold seem to be?
It depends on what the number of iteration is, but with 100 iterations, from step 0.1 to 6.5 it finds a solution close to 0. But right after 6.6, it starts to diverge and for any value after 6.7 it diverges. The treshold seems to be around 6.6.

1.2 WU2

Come up with a non-convex univariate optimization problem. Plot the function you're trying to minimize and show two runs of gd, one where it gets caught in a local minimum and one where it manages to make it to a global minimum. (Use different starting points to accomplish this.)
Using $f(x) = sin(\pi x) + x^2/2$, $f'(x) = \pi cos(\pi x) + x$, and 10 iterations, the global minimum happens at $x \approx -0.45385$. If we start at 0, we can find the global minimum, but if we start at 1, gd gets caught in a local minimum 1.357

```
>>> gd.gd(f, derF, 0, 10, 0.2)
(-0.45385351939658519, array([ 0.         , -0.72244765, -0.84494752, -0.88472033, -0
       -0.88651822, -0.88651823, -0.88651823, -0.88651823, -0.88651823,
       -0.88651823]))
>>> gd.gd(f, derF, 1, 10, 0.2)
(1.3577434052579487, array([  1.22464680e-16,   4.52961014e-02,   2.50055080e-02,
         2.00175811e-02,   1.99477186e-02,   1.99477147e-02,
         1.99477147e-02,   1.99477147e-02,   1.99477147e-02,
         1.99477147e-02,   1.99477147e-02]))
```

1.3 WU3

Why does the logistic classifier produce much larger weights than the others, even though they all get basically the same classification performance?
Don't have to answer this.

2 Warm Up with ML Tools

2.1 WU4

What are the five features with largest positive weight and what are the five features with largest negative weight? Do these seem "right" based on the task?

1. motif -1.21427941322326660156
2. window -1.15418136119842529297
3. server -0.94833439588546752930
4. list -0.89077484607696533203
5. x -0.86568439006805419922

1. graphics 1.09135174751281738281
2. images 0.72243607044219970703
3. image 0.72001230716705322266
4. card 0.71239823102951049805
5. xx 0.69124054908752441406

2.2 WU5

Draw the tree. How do the selected features compare to the features from the logistic regression model? Which features seem "better" and why? If you use a depth 10 tree, how well do you do on test data?
TODO: add the image of the tree later The selected features include four of the best/worst features from the logistic regression ("graphics", "motif", "window", and "x").

Features can be considered "better" when the ratio bewteen class 1 and class 0 at the leaf is higher and when the frequency of the number of instances that got to that leaf is high. In this respect, "motif" and "vga" are the better features. "motif" has high frequency (total 583 instances fall have no motif) and high ratio (about 80% of those without "motif" are in class 1). 395 instances fall in "vga", and 84% of those with no "vga" are in class 0. If

they have “vga”, 100% of them are in class 1 (although this may not be the best indicator because only 13 data falls into yes “vga”.) In this respect, “usr” is one of the worst features in this tree because the relative frequency is low, adn when there is no “usr”, the ratio of class1 to class0 is 6:4.

If I use a depth 10 tree, I get a test error of 20.53%, (0.463% improvement).

2.3 WU6

2.4 WU7

3 Reductions for Multiclass Classification

3.1 WU8

3.2 WU9

3.3 Ranking or Collective Classification

3.4 WU10a

3.5 WU10b