Computational Advertising | Alex Smola
Lecture: April 22, 2009 | http://alex.smola.org
UC Santa Cruz | Yahoo! Labs, Santa Clara

Problem Set — Machine Learning

Problem 1 (Data Set Preparation (10 points))
Download sample machine learning datasets from the UCI Irvine Repository at http://archive.ics.uci.edu/ml/.

1. *Obtain 5 datasets for binary classification (hint: use real-valued datasets).*
2. *Import them into MATLAB, Python with NumPy, R, or any other numerical analysis platform of your choice.*
3. *Create 2 copies of each dataset*
   - *Original dataset*
   - *Rescale data to* 0 *mean and unit variance coordinate-wise*

Provide the (pseudo)-code for these routines.

Problem 2 (Perceptron (20 points))

1. *Implement the basic Perceptron algorithm for binary classification.*
2. *Split each dataset (all 3 copies) 80 / 20 into a training and a testing dataset.*
3. *Compute the average classification error on training and test set after running the Perceptron once on the training set, and after 10 passes over the training set.*
4. *Incorporate a decreasing learning rate* $\eta_t = \frac{c}{\sqrt{t+t_0}}$ *(e.g.* $c = 0.1$ *and* $t_0 = 1000$*) and repeat the experiments.*

Provide the (pseudo)-code and errors for these routines. What happens if you change c *and* t_0*?*

Problem 3 (Rotation Invariance (10 points))
Prove that the stochastic gradient descent algorithms are invariant under rotation. That is, show that if we replace x *by a rotated copy,* Ux *where* $U^\top U = \mathbf{1}$*, then the classification / regression results of the algorithm remain unchanged.*

Problem 4 (Game Show (10 points))
Assume that in a TV show the candidate is given the choice between three doors. Behind two of the doors there is a pencil and behind one there is the grand prize, a car. The candidate chooses one door. After that, the showmaster opens another door behind which there is a pencil. Should the candidate switch doors after that? What is the probability of winning the car?

Problem 5 (Robustness — bonus (10 points))
Show that for the problem of quadratic regression, that is, find a function $f(x) = \langle x, w \rangle$ *such that the loss* $\sum_i (y_i - f(x_i))^2$ *is minimized, a single choice of* y_i *suffices to change the* f *by an arbitrary amount.*

Show that if we replace the loss by $\sum_i |y_i - f(x_i)|$ *the change in* f *is bounded if we change* y_i *arbitrarily. Hint: use the first-order optimality conditions of the optimization problem.*