Machine Learning 2
Exercise Sheet 8

Prof. Dr. Dr. Lars Schmidt-Thieme, Brad Baker
Information Systems and Machine Learning Lab
University of Hildesheim

June 21st, 2017

Submission until June 27th, 8:00 AM by learnweb.
Please put your name in all filenames.
Non-pdf submissions for non-programming exercises will not be graded.

Exercise 15: Least Angle Regression (10 points)

a) (3 points) Briefly describe the Homotopy Method at work in Least Angle Regression and its relation to Forward-Search for var. Show how this idea allows us to write L1-Regularized Regression as an optimization problem with one constraint.

b) (7 points) Suppose you are performing Least Angle Regression and are at update step t. At this step, you are working with the following predictors, X and estimated targets $\hat{y}^{(t)}$, as compared with the ground-truth y:

x_1	x_2	x_3	$\hat{y}^{(t)}$	y
4.5	2.3	1.3	0.275	0.3
5.2	3.4	1.4	0.350	0.2
5.7	4.4	1.5	0.420	0.4
4.6	3.6	1.0	0.280	0.2
5.0	3.3	1.4	0.345	0.2

Table 1: Predictors X, estimated target $\hat{y}^{(t)}$, and GT y

The parameters at step t are given as

$$\beta^{(t)} := \begin{pmatrix} -0.1 & 0.0 & 0.05 & 0.2 \end{pmatrix}$$

- Choose the predictors with the largest absolute correlation with the residuum, and regress these predictors on the residuum to obtain an update for these parameters $\hat{\gamma}^{(t)}$.
- For a step-length of $\alpha = 1$ compute $\hat{y}^{(t+1)}$, compare the change in the Residual Sum of Squares as $\Delta_{RSS} = RSS(y, \hat{y}^{(t+1)}) - RSS(y, \hat{y}^{(t)})$.
- Finally, compute the updated correlation $C_m^{(t+1)}$ for your new predictions, and list the variables in the new active set.

NOTE: You may use a calculator or software to do the arithmetic for this exercise, but at this point, it's a good idea to try to do the algebra mostly by hand, as practice for the exam. A maximum of one point will be taken off for small arithmetic mistakes, but more points will be lost for algebraic errors, or major arithmetic mistakes.

Exercise 16: Least Angle Regression with Stepsize Adjustment

a) (3 points) After performing a step in LARS, the updated correlation in step $t+1$ can be given as

$$C_m^{(t+1)} = C_m^{(t)} - \alpha X_{\cdot,m}^T X^{(t)} \hat{\gamma}^{(t)}.$$

The idea behind LARS with adjusted stepsize is to choose the step size in this update, α, such that we reduce our active predictors until we meet an inactive predictor at the same (maximum) residuum correlation. Show how this intuition provides us with a formula for updating α by using the correlations from the previous iteration. This formula is given on slide 7 of the lecture.

b) (7 points) In the previous exercise, we chose a step-size arbitrarily as $\alpha^{(t)} = 1$

- Using the formula you illustrated in part a), recompute the step-size $\alpha^{(t)}$, and then compute the updated parameters $\beta^{(t+1)}$.
- Compute the updated correlation $C_m^{(t+1)}$ for your new predictions, and list the variables in the new active set.
- Given the variables in the remaining inactive set and their correlations which you have just computed, hypothesize which variables will get added to the active set at the end of the next iteration. How will this choice affect the update of α in the next iteration?

Bonus: Least Angle Regression in R (10 points)

For this exercise, it is recommended you use R because there are libraries to easily compute and plot LAR; however, you can also use python if you want to figure out the python equivalents on your own.

If you are using R, you should use the LARS package for this problem `https://cran.r-project.org/web/packages/lars/lars.pdf`

You will be using the IRIS dataset for this problem, which is one of the default datasets loaded into R.

- First, break the dataset into three subsets based on the species of IRIS.
- Each subset now has four variables - for each subset, explore the effectiveness of the Petal Length and Width as Predictors for 1) Sepal Length and 2) Sepal Width, by performing LAR and plotting both the change of the coefficients and the prediction error for each task. You should in total have 12 plots - 2 different plots for the 2 different prediction-targets for the 3 different subsets.
- Briefly summarize your findings w.r.t each of the tasks for each of the subsets.