

CAP6676 Knowledge Representation, Fall 2015

Instructor: Guo-Jun Qi. Due: November 3rd, 2015.

Machine Problem 2: Logistic Regression: Learning and Inference

In this machine problem, you will implement the (stochastic) gradient ascent algorithm to learn the parameters for logistic regression model.

Step 1: Use the same dataset in MP2 for the experiments. You will need to compare the results between LR model and decision tree model.

Step 2: Implement the learning algorithm for LR model with batch-mode gradient ascent algorithms. You should NOT use any off-the-shelf optimization library here. You need to derive the derivative of loglikelihood function to the model parameters over the entire training set, and update the model parameters iteratively. Choose a stop criterion to decide when to terminate the iteration.

For example, you can stop the algorithm when

- the increment of loglikelihood between two successive iterations is smaller than a threshold, OR
- the absolute value of the derivatives is smaller than a threshold, OR
- a maximum number of iterations has been reached.

You should monitor and plot the curve of the objective loglikelihood over iterations, and report your observation in your report. Also, plot how the training and test error evolve over iterations, and report if you observe any overfitting phenomenon (i.e., decreasing training error but increasing test error).

Step 3: Run stochastic gradient ascent algorithms. Choose a proper size of randomly chosen subset. Repeat the process in Step 4. In addition, report how the size of subset in each iteration affects the result.

Step 4: Add the L_2 regularizer $-\frac{1}{2\sigma^2}||w||^2$ to the likelihood function, and maximize this regularized loglikelihood function by gradient and stochastic gradient ascent methods. Repeat Step 2 and Step 3.

In addition, tune the parameter σ on a separate validation set or by a cross-validation process. Report if this regularization mechanism can reduce the overfitting effect.

Step 5: Write the report following the above steps. Submit your report to cap6676ucf@gmail.com by November 3rd. Remember to compare the LR result with that of decision tree, report see which is better on your dataset. Explain why?

Note: In case that your chosen dataset contains categorical values of attributes, you can convert them into a binary vector indicating the categorical values. For example, for an attribute with K possible categories, construct a K-dimensional binary vector with the entry corresponding to the category set to one, and the other entries set to zero.