Machine Learning Module

Week 4

Laboratory Exercise, Week 4

Probabilistic Classification Methods

Mark Girolami
girolami@dcs.gla.ac.uk
Department of Computing Science
University of Glasgow

January 13, 2006

1 Laboratory Exercise

In this session we will now use our knowledge of probabilistic linear modeling to look at a richer class of function approximators than the simple polynomial-based methods.

1.1 Discriminative Classifier

Download the Matlab and data files to your machine from the Week 4 directory. File rip_dat_tr.txt contains the training data where the feature vectors for each sample are listed in the first two columns and the class label is given in the third column. The file rip_dat_te.txt contains the corresponding test data.

The logistic regression classifier will use a simple polynomial basis expansion of the two-dimensional data vectors and the object of this laboratory is to investigate how the classification performance is affected by the polynomial order employed. The other hyper-parameter which has an effect on the model performance is the prior variance as we saw in the previous laboratory session.

Compute the the miss-classification rate on the test set and the associated test set predictive likelihood for a range of polynomial model orders of $K = 1 \cdots 20$ and prior variances α ranging from 1 to 100.

You should be able to produce a series of plots similar to the one below for $K = 3$. From this chart we can see that as we move away from the maximum likelihood solution (by making the prior variance smaller (and hence increasing the amount of regularisation of the solution) the training error increases whilst the test error actually decreases.

Do we see similar behaviour if we monitor the log-likelihoods computed on the train and test data? What does the log-likelihood computed on test data tell us that the number of miss-classified test points does not?

Is there any significant difference in classification accuracy when using the MAP estimate compared to approximate posterior averaging?

2 Generative Classifier

The file 20news_w100.mat contains around 16,000 documents, represented as binary vector models, from four classes of Newsgroup[1] with a dictionary size of 100 words (miniscule in IR terms). You will notice that the matrix of the documents and words is in sparse format.

A Matlab script naive_bayes_binary.m is available which demonstrates how a Naive Bayes classifier performs when classifying over four classes of document. How would the logistic regression method perform when attempting to separate Classes 1 & 2 ? how does this compare to the Naive Bayes Classifier. Discuss.

[1]Data prepared and made available by Sam Roweis, University of Toronto

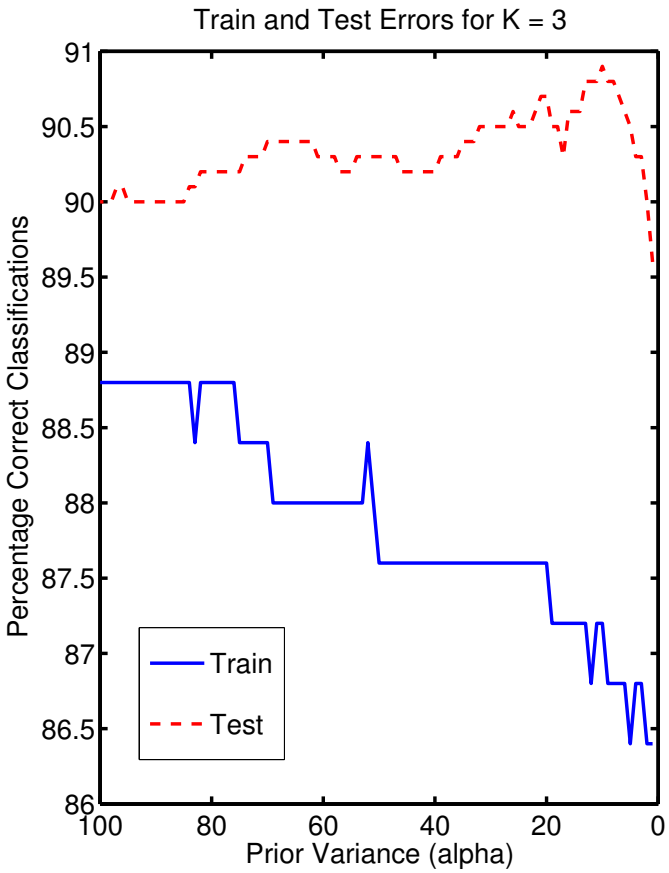

Figure 1: The plots show how train and test errors vary with α the amount of regularisation imposed for a cubic order model.