

# Homework 8

Seongjin Bien, Mohit Shrestha  
Machine Learning - Spring 2018

April 15, 2018

For this assignment, we trained a sequence of different classifiers with increasing model flexibility, with a varying number of feature vectors. For each of our classifiers, we followed the 6-step procedure documented in the homework sheet. The features we chose were all obtained via PCA, extracting the first  $k$  columns from the PCA matrix. We varied the linear regression results from  $k = 1$  to  $k = 240$ . The results of this can be seen in Figure 1:

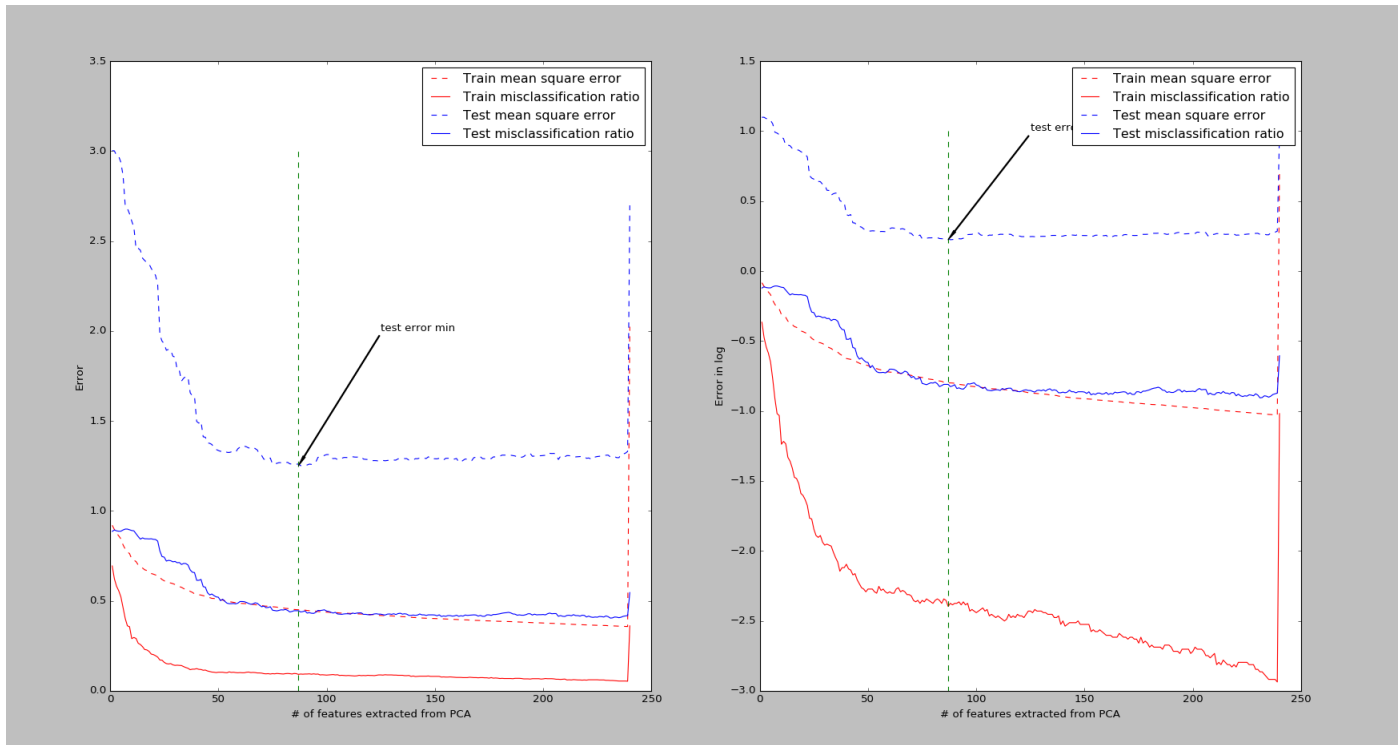


Figure 1. Left: linear scaling, Right: logarithmic scaling

As expected, the  $MSE_{Test}$ ,  $MC_{Test}$ ,  $MSE_{Train}$  and  $MC_{Train}$  all decrease logarithmically with increasing number of features  $k$ . Both  $MSE_{Train}$  and  $MC_{Train}$  decrease as expected with increasing number of  $k$ . However, this trend only continues until  $k = 87$ , at which the minimum  $MSE_{Test}$  is reached. This is more noticeable in the log plot. The value shows

a minor increase afterwards, which implies overfitting, but this is not very significant. We assume that this is due to us using PCA and linear regression together, which results in affine linear map from beginning to the end. This results in a less-than-ideal result, as shown here in the miniscule increase in  $MSE_{Test}$  and the fact that the convergence rate is much slower than the textbook example of  $k$  around 40, meaning more features are needed to reach the same point.  $MC_{Test}$  shows a much more inconsistent behaviour and greater jitter, as it was only implicitly targeted by the algorithm.