02457 Non-Linear Signal Processing: Exercise 3

Checkpoint 3.1:

The program **main3a.m** is used to create a training-set with a 2-dimensional input variable and a 1-dimensional output variable. The estimated weight vector is compared with the true one and the dependence on both the noise level and number of points in the training-set is investigated.

A good fit requires enough data points. Good fit with 3 points, but need many more, the error converges to noise level squared.

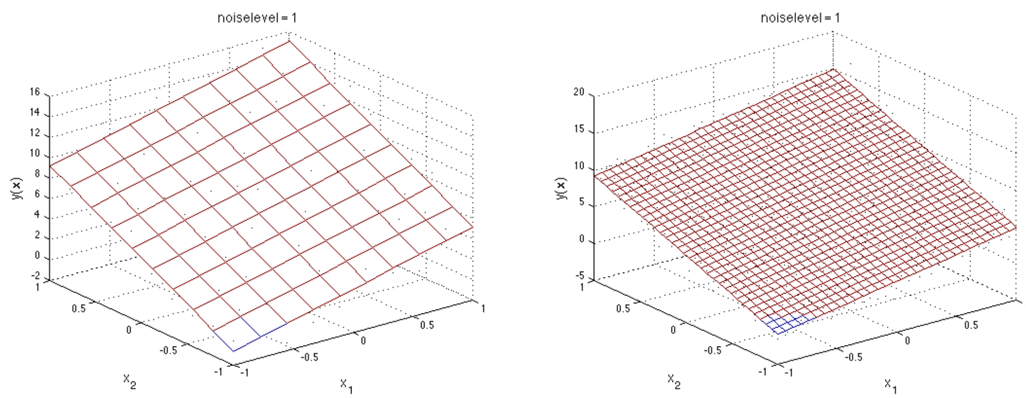

Figure 1: **Left:** 100 training points. **Right:** 1000 training points.

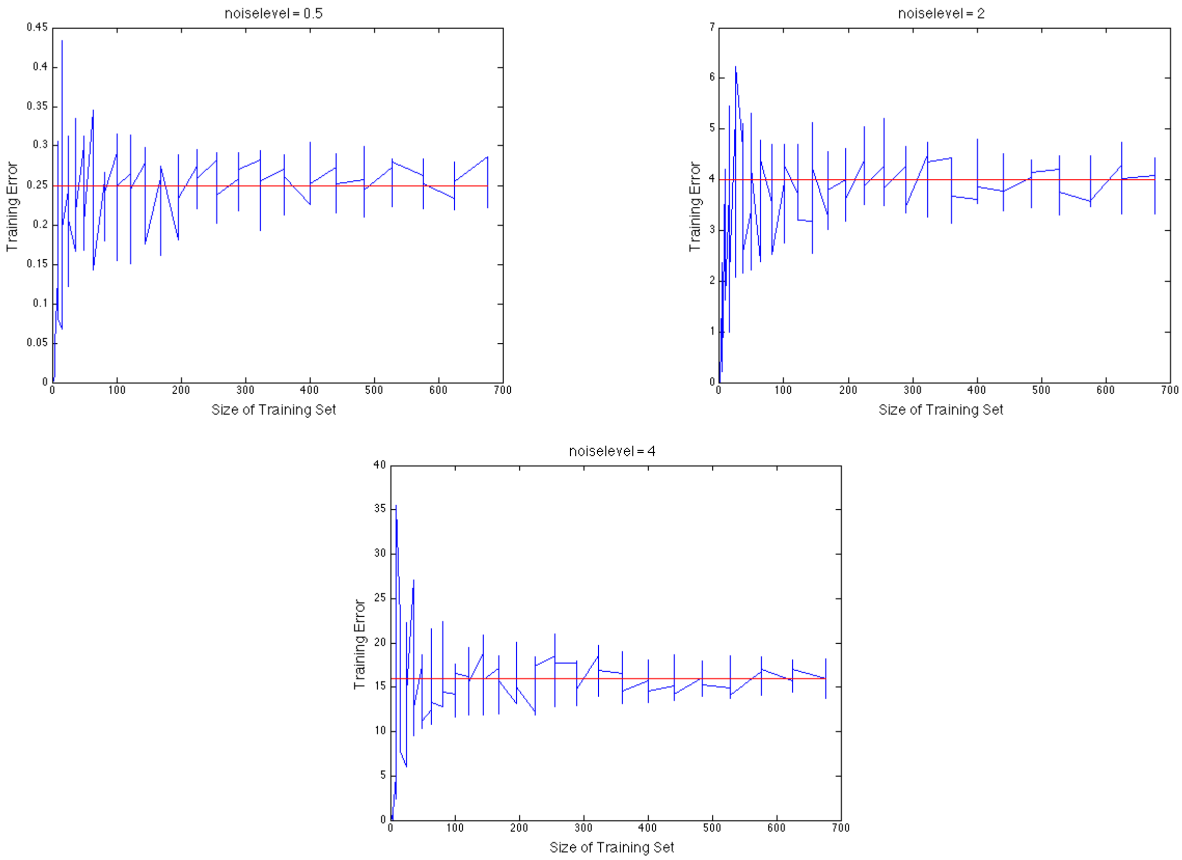

Figure 2: The training error as a function of training set size for different noise levels.

N	noise=1, $\lvert w_{true} - w_{est} \rvert$	noise = 1, $\lvert w_{true} - w_{est} \rvert / w_{true}$	noise = 2, $\lvert w_{true} - w_{est} \rvert / w_{true}$
10	[0.1254 0.0080 0.1130]	[0.0054 0.0039 0.0071]	[0.0158 0.0045 0.0470]
50	[0.0560 0.0761 0.0732]	[0.0043 0.0030 0.0054]	[0.0032 0.0611 0.0397]
100	[0.0375 0.0118 0.0353]	[0.0054 0.0039 0.0071]	[0.0121 0.0156 0.0114]
500	[0.0034 0.0072 0.0049]	[0.0007 0.0027 0.0070]	[0.0042 0.0231 0.0075]
1000	[0.0004 0.0046 0.0045]	[0.0005 0.0075 0.0012]	[0.0044 0.0051 0.0040]
5000	[0.0086 0.0078 0.0077]	[0.0004 0.0033 0.0004]	[0.0005 0.0030 0.0039]

We see that the estimate of the weights get increasingly better for larger training set size, and that it gets worse for higher noise levels.

Checkpoint 3.2:

We have a data set of size 280, and when we choose to regress over d dimensions (look d years back to predict the next year), there are a remaining 280 - d dimensions that we are trying to predict.

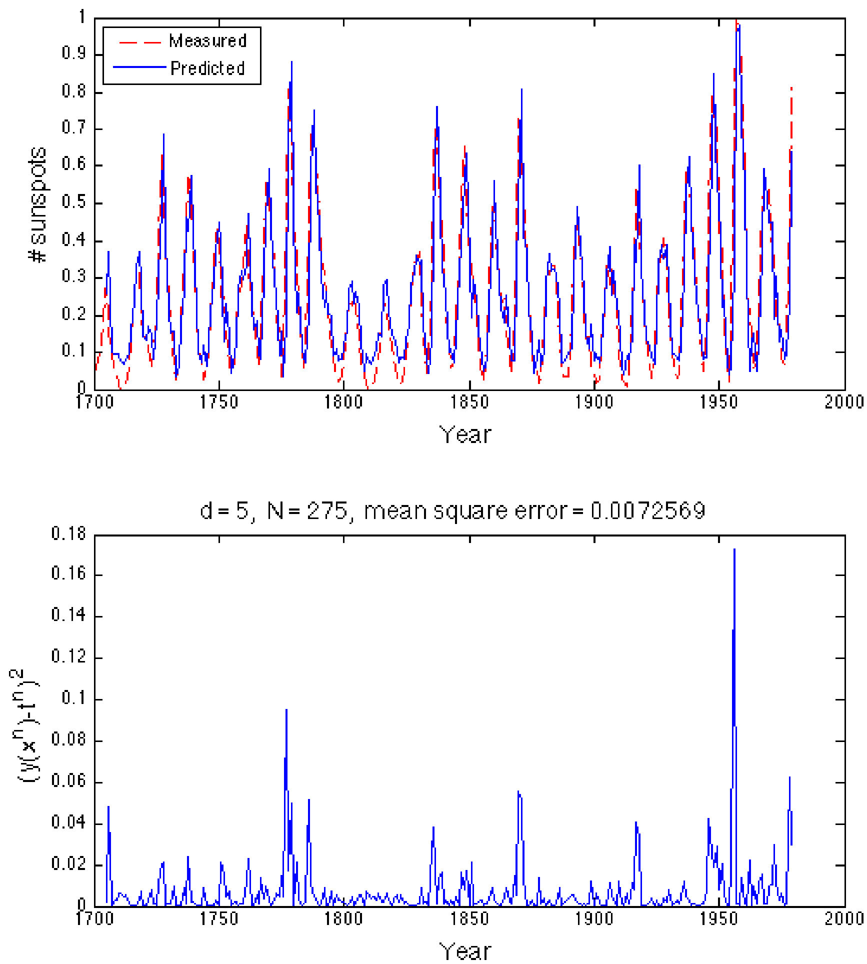

Figure 3: Example of a prediction together with the data.

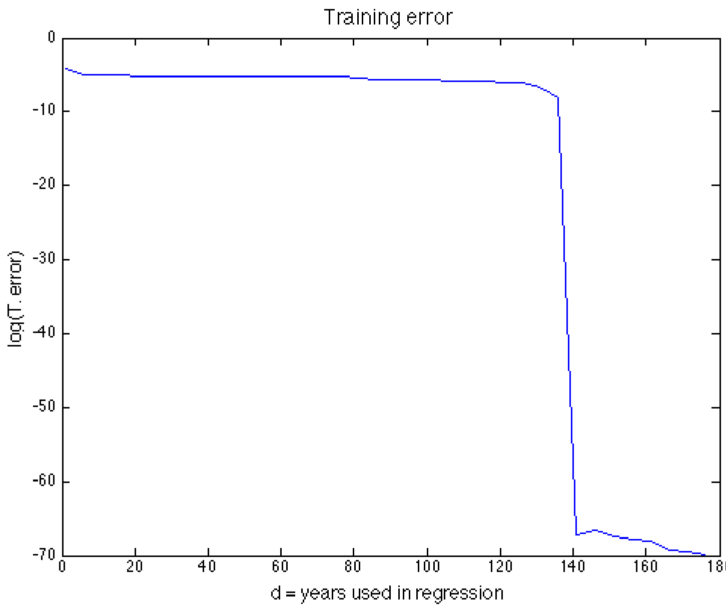

Figure 4: Training error as a function of d.

Checkpoint 3.3

The program **main3c.m** is used to find the direction maximizing class separation for a two-class problem. The projection of the data-set onto one-dimension with the projections found using eigenvector transformation is compared to the direction found using Fisher's Linear Discriminant.

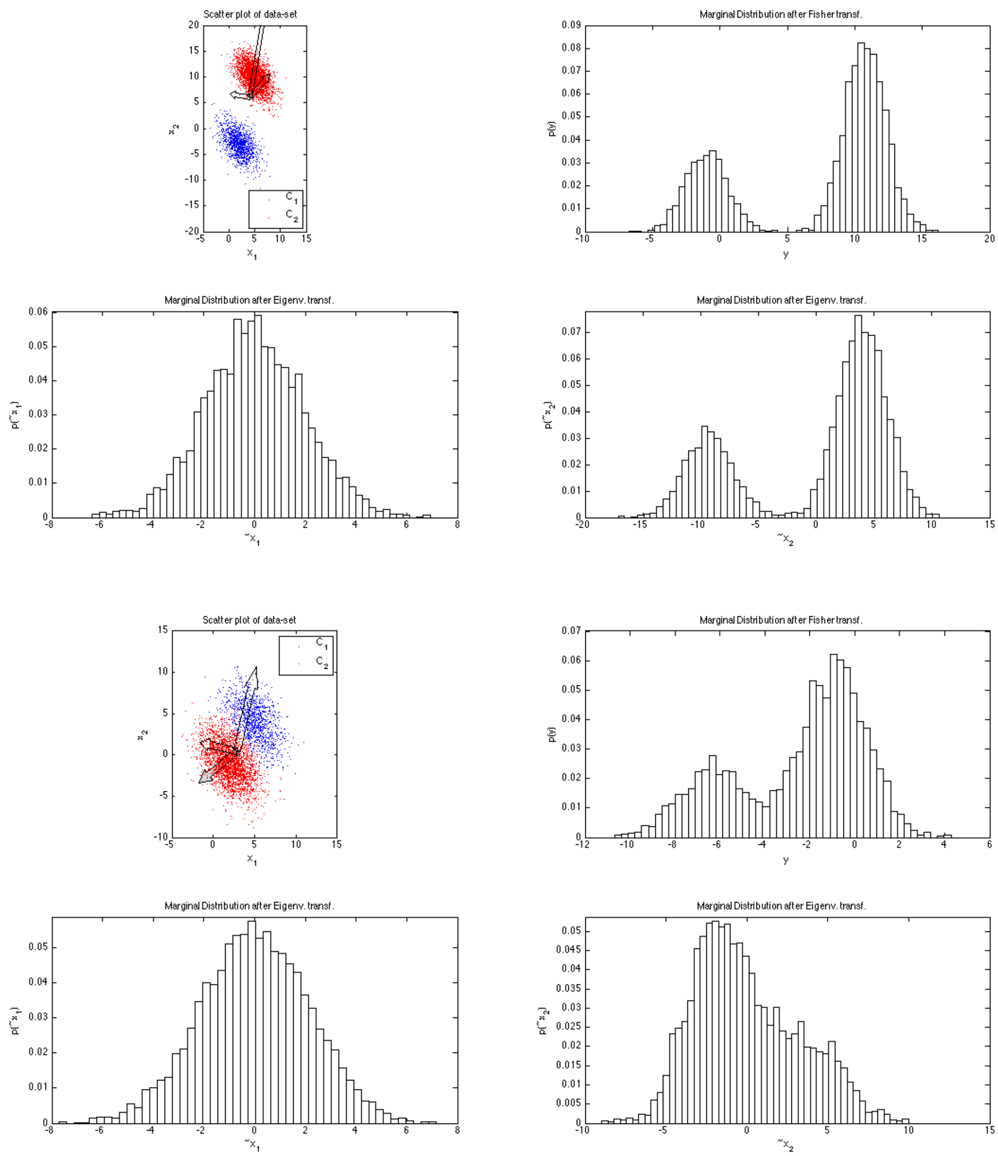

Figure 5: **Top:** Example where both transformations are good. **Bottom:** Fischer is best.

Fisher's linear discriminant can be found by maximizing Fisher's criterion in an iterative procedure:

$$j(w) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \tag{1}$$

where m_1, m_2 are the means of the two classes and s_1^2, s_2^2 are their variances.