MLAP Open Assessment A

Y6186228

14th May 2014

1 Linear regression and Logistic regression

1.1 Task 1

In order to experiment with linear regression I have chosen to use the following 8 features, referred to in equations as a through h for brevity:

a	stock volume of previous day
b	difference between the previous two days' stock volumes
c	mean of stock volumes from previous ten days
d	standard deviation of stock volumes from previous ten days
e	stock price of previous day
f	difference between the previous two days' stock prices
g	mean of stock prices from previous ten days
h	standard deviation of stock prices from previous ten days

Further, the elements of a vector θ represent the coefficients of a regression function, with θ_0 always representing the constant term. Hence, a regression function might look as follows:

$$f(\theta) = \theta_0 + \theta_1 a + \theta_2 a^2$$

Figure 1 shows the Mean Squared Errors (MSEs) obtained in my initial phase of experimentation with the chosen features, shown to three significant figures. In this phase of experimentation, I evaluated the performance of each feature used on its own in first- and second-order polynomials.

It is clear from these results that the best performances come when using the features that relate to stock price as opposed to stock value ($e - h$). However, feature f, the difference between the last two days' stock prices does not appear to perform very well. Feature e does not perform well as a first-order polynomial, but is exceptionally good as a second-order polynomial.

My next phase of experimentation is to take the high-performing features and try using them on their own in third-order polynomial functions. Following this, I will experiment with combining the better performing features to see what improvements can be made. The results of the initial third-order polynomial experiments are shown in figure 2. Out of interest, I have chosen to try feature c as a third-order polynomial as it was the best performing of the stock volume-related features.

As seen by the results, each feature tested in third-order polynomials had very similar results to the second-order polynomial tests. Going forward, I have chosen to try a function combining e and g both as second-order polynomials to see if they perform well as a pair. I am also interested to to see if the recent change in stock volume (b) combined with the mean of the last ten days' stock prices (g) gives an indication of the next stock price. Further, combining a, b, d and g all together may give good results. The MSEs obtained for these tests are shown in figure 3 (note the change of scale for this chart).

From these results we can see

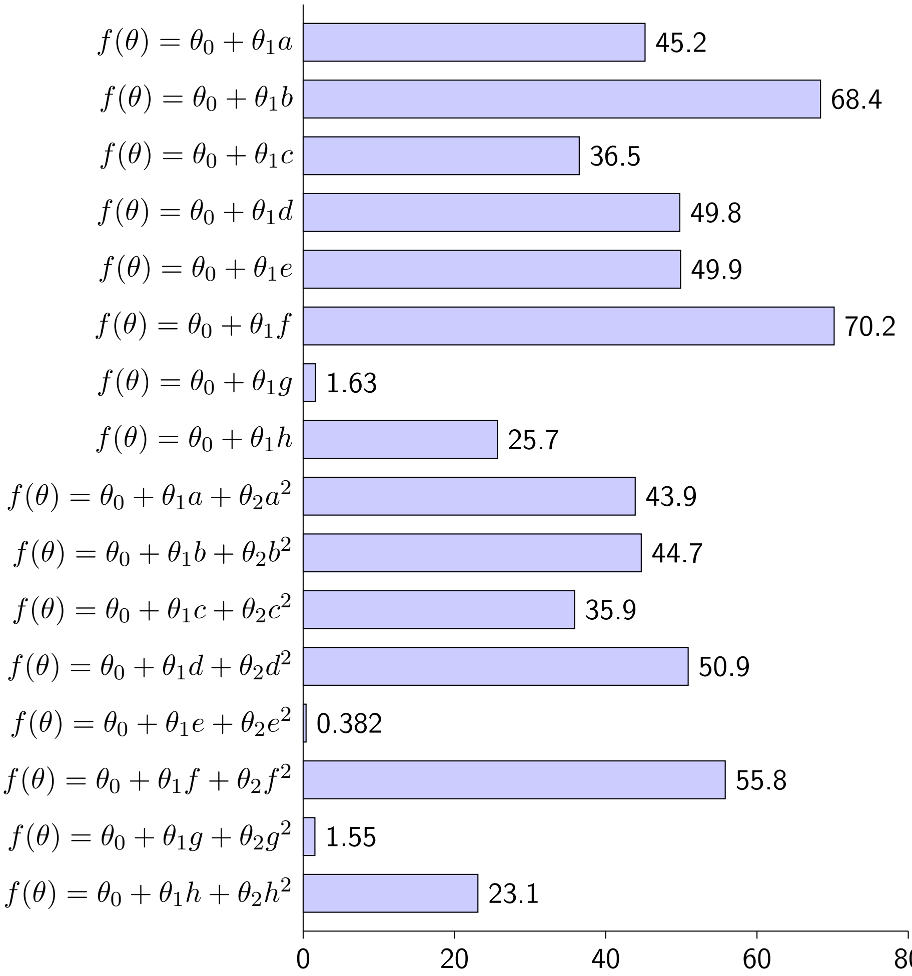

Figure 1: MSEs obtained using each of the features on their own in first- and second-order polynomials

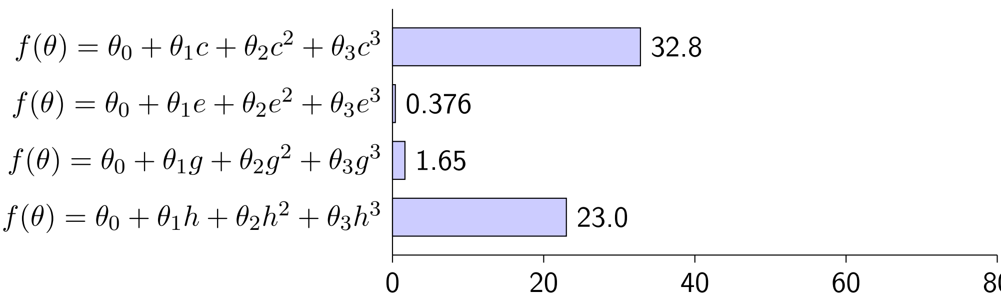

Figure 2: MSEs obtained using selected features on their own in three-order polynomials

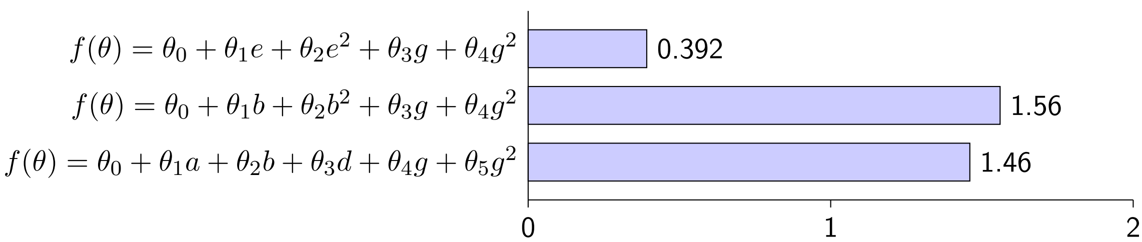

Figure 3: MSEs obtained when combining features into more complex polynomials

1.2 Task 2

1.3 Task 3

2 Bayesian networks

2.1 Task 4

2.2 Task 5