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:

$$f(\theta) = \theta_0 + \theta_1 a + \theta_2 a^2 + \theta_3 b$$

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 that combining e and g does not really improve on the performance achieved when using e alone. Also, combining g variously with a, b and d does not improve on the performance of using g alone.

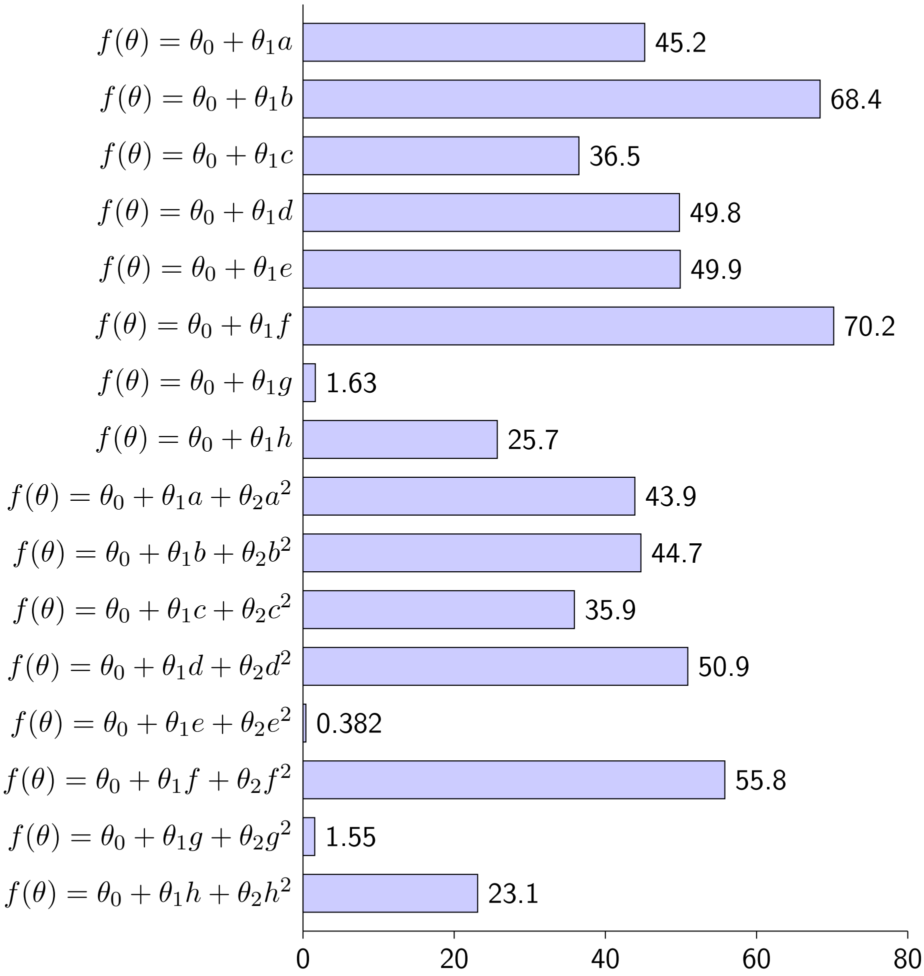

Figure 1: MSEs obtained using each of the features on their own in first- and second-order polynomials

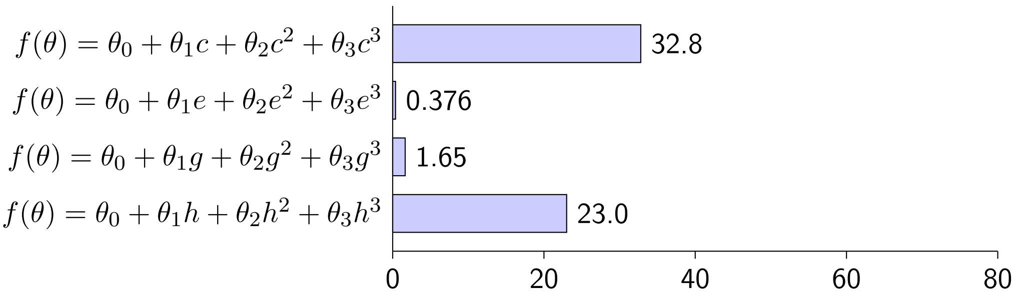

Figure 2: MSEs obtained using selected features on their own in three-order polynomials

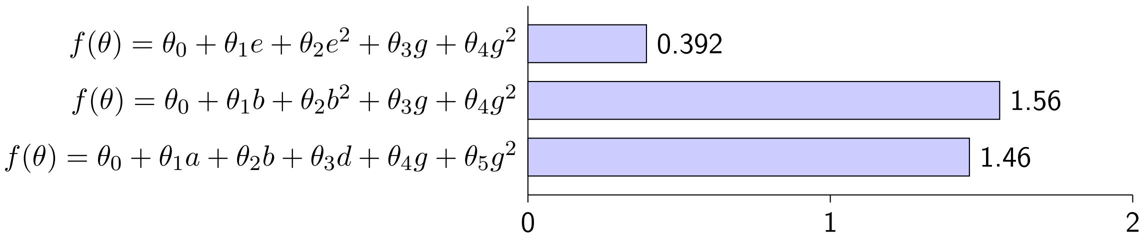

Figure 3: MSEs obtained when combining features into more complex polynomials

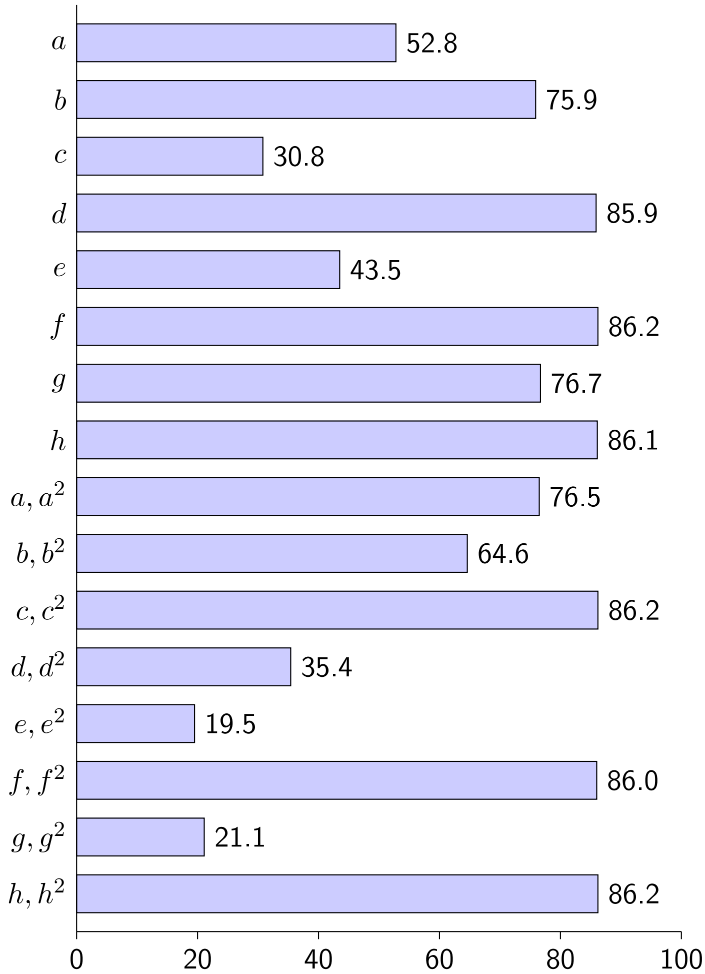

Figure 4: % accuracies obtained using each of the features on their own in first- and second-order polynomials

1.2 Task 2

The same features, represented as a through h, are used in this task. For simplicity, I will simply list the features and powers used in each test (for example, a, a^2, b). The constant term is implicitly used in every case.

Again, I started by evaluating the use of each feature on its own in first- and second-order polynomials.

Figure 4 shows the percentage accuracy achieved by each of the tests using each feature individually in first- and second-order polynomials. As can be seen, the high performing features (f, h) did not benefit from being included in a squared form. Some of the features were better used in their second-order polynomials than in their first-order ones (e.g. a) while others were significantly worse (e.g. b).

I now will test the performance of using different combinations of features. One such combination is f, h, to see if combining them improves on their individual successes. Also, the pairs e, h and g, h, since each pair can give an idea for the general spread of recent stock price data in terms of percentage. The effect of relationship between c, g is also interesting to test, as well as the quadruple of a, c, e, g, since it includes information about the relationship between the last day's data and the mean data from the last 10 days. I will also see what happens if I combine all four of the features that achieved scores in the region of 86% individually (a, d, f, h). In each test, each feature will be raised to the power at which it performed best from the previous tests.

As with task 1, it is seen from the results that combining features (figure 5) did not really improve the performance. Firstly, c, c^2, d, f, h gave worse results than all of its constituents' individual results. As well as this, f and h perform as well individually as when combined together. The e, h result is worse than the independent results of h, though better than both independent results for e. The g, h result is worse than the best independent performances of g and h. c, c^2, g performed poorly. a, a^2, c, c^2, e, g pretty much matches but does not outperform the best performances of 86% from the previous tests, and this is also true

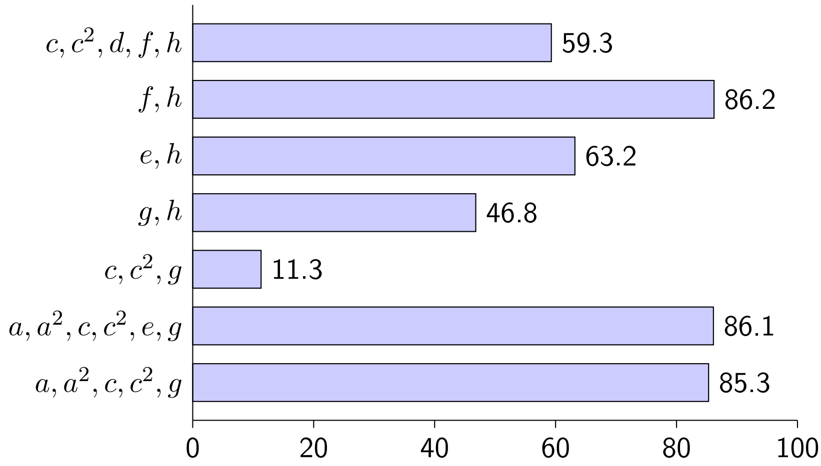

Figure 5: % accuracies obtained when combining features into more complex polynomials

for a, a^2, c, c^2, g.

It is interesting to see from these results that the accuracies seem to max out at about 86%. From looking at the actual classifications given by the higher-performing functions I have found that this actually occurs when every row of data is predicted as being in class 0. Some of the functions giving lower accuracies give classification distributions that appear more realistic, but since the evaluated accuracies are lower we must assume that it is the wrong data rows which are being predicted as classes other than 0. Therefore, I am lead to conclude that the best policy found from these experiments is to predict all data rows as being in class 0. These experiments have not been totally comprehensive, however, so it is possible that a function might exist that can achieve a better accuracy (either by including higher polynomials or different combinations of features, or both).

1.3 Task 3

2 Bayesian networks

2.1 Task 5 Question 1

A conditional probability $P(A|B)$ could be estimated by selecting all of the instantiations that satisfy the condition B and calculating the proportion of those for which A is true.

For example, if we want to estimate $P(7 = 1|6 = 1)$ (The probability that node 7 is set to 1 given that node 6 is set to 1) then we produce a sample and select from it all of the instantiations in which node 6 is set to 1. Figure 6 shows a full sample of 10 instantiations, with the selected instantiations marked using an asterisk (*).

```
[1, 0, 0, 0, 0, 0, 1, 1] *
[0, 0, 1, 0, 0, 0, 1, 1] *
[1, 0, 1, 0, 0, 0, 1, 1] *
[0, 0, 0, 0, 0, 0, 1, 0] *
[1, 0, 0, 0, 0, 0, 1, 0] *
[1, 0, 1, 0, 0, 0, 1, 0] *
[0, 0, 1, 0, 0, 0, 1, 1] *
[0, 0, 0, 0, 0, 0, 0, 0]
[0, 0, 0, 0, 0, 0, 1, 0] *
[1, 0, 1, 0, 0, 0, 1, 1] *
```

Figure 6: A sample of size 10, with the instantiations where node 6 is set to 1 marked by asterisks

From the selection, we find out how many instantiations also have node 7 set to 1. Figure 7 shows just the 9 selected rows, and marks the rows in which node 7 is set to 1 with a plus sign (+).

```
[1, 0, 0, 0, 0, 0, 1, 1] +
[0, 0, 1, 0, 0, 0, 1, 1] +
[1, 0, 1, 0, 0, 0, 1, 1] +
[0, 0, 0, 0, 0, 0, 1, 0]
[1, 0, 0, 0, 0, 0, 1, 0]
[1, 0, 1, 0, 0, 0, 1, 0]
[0, 0, 1, 0, 0, 0, 1, 1] +
[0, 0, 0, 0, 0, 0, 1, 0]
[1, 0, 1, 0, 0, 0, 1, 1] +
```

Figure 7: The selected rows from figure 6, with the instantiations where node 7 is set to 1 marked by plus signs

We can see that, of the 9 instantiations that satisfy $6 = 1$, 5 also satisfy $7 = 1$. We therefore estimate that $P(7 = 1|6 = 1) = \frac{5}{9}$.

2.2 Task 5 Question 2

I have yet to answer this question.

2.3 Task 5 Question 3

I will use the same example in this question as in question 1 (find $P(7 = 1|6 = 1)$). From the original data provided in bndata.csv, we know that the probability is $3688/8887 = 0.415$ (3dp). Using the method described above, I will estimate the probability based on 100, 1000, 5000 and 10,000 samples and record them in the table below. The final column shows the percentage error of the estimation.

Samples	Number where $6 = 1$	Number where $7 = 1$	Estimated probability (5dp)	Error (3sf)
100	87	28	0.32184	22.4
1000	885	360	0.40678	0.0198
5000	4450	1834	0.41213	0.00692
10,000	8888	3691	0.41528	0.000675

Here, the accuracy of the estimation is shown to get gradually better and better as the number of samples increases from 100 to 10,000.