Assignment 1, COMP4702

Roy Portas

March 26, 2017

Question 1.2

The first column of the data is the date, the second column appears to be a unique ID for each entry.

The third column contains numbers between 25 and 30, this is possibly a temperature in degrees. The values also change gradually which seems correct for the given time intervals

The fourth column contains numbers between 26 and around 50000. If this is plotted against the ID field, it produces a line.

The fifth column contains numbers between 7.3 and 8.3, with a mean of 7.846 and a standard StdDev of 0.142. This suggests that the value doesn't change much.

It could possibly be weather data, containing temperature, humidity, etc.

Question 1.6

```
% in is the input array
% n is the group size
function out = q6(in, n)

    out = [];

    chunks = length(in)/n;

    for i = 1:chunks
        end_ind = length(in) - i * n + n;

        start_ind = end_ind - n + 1;
        start_ind = max([1, start_ind]);

        temp = in(start_ind:end_ind);

        out = [out, temp];
    end

end
```

Question 2.1

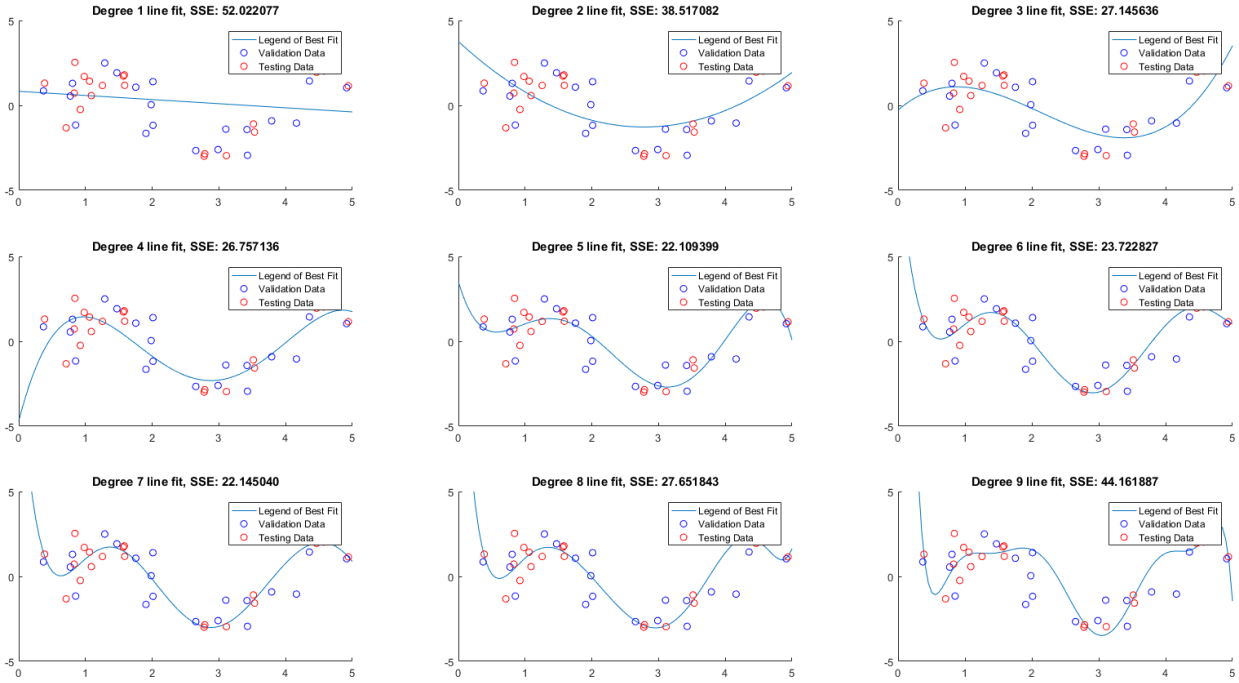

Figure 1: Lines of best fit

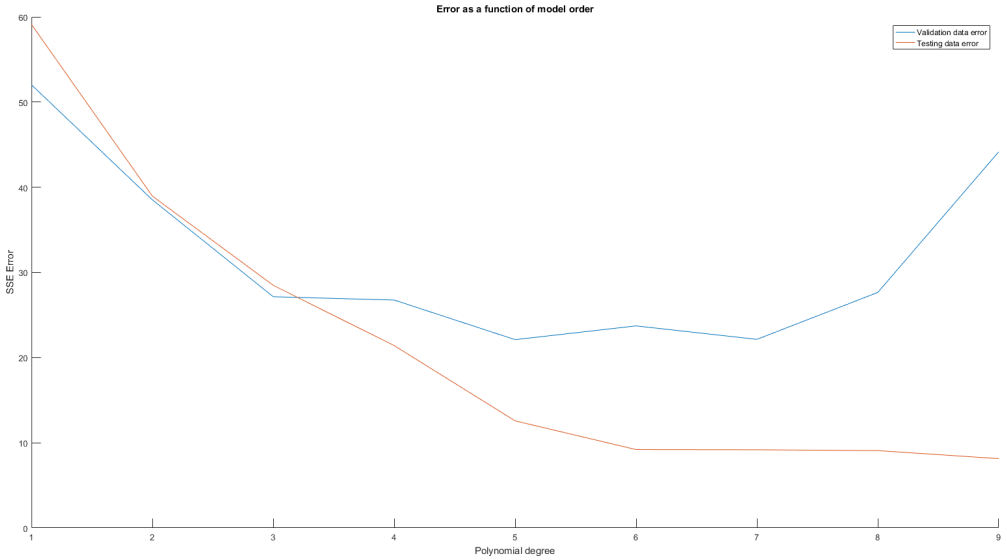

Figure 2: Error vs Polynomial Degree

Question 2.4

```
function q4(data, class)
    class_names = unique(class);
    x_values = 1:length(class);

    class1 = zeros(1, length(class));
    class2 = zeros(1, length(class));

    for i = x_values
        if strcmp(class{i}, class_names{1})
            class1(i) = 1;
        else
            class2(i) = 1;
        end
    end

    % Verify the classes are correct
    % figure;
    % hold on;
    % scatter(1:length(class1), class1);
    % scatter(1:length(class2), class2);
    % hold off;

    estimate_range = 1:0.1:8;

    class1_data = data;
    class1_data(class2 == 1) = NaN;

    class1_mle = mle(class1_data);
    class1_pdf = normpdf(estimate_range, class1_mle(1), class1_mle
        (2));

    class2_data = data;
    class2_data(class1 == 1) = NaN;

    class2_mle = mle(class2_data);
    class2_pdf = normpdf(estimate_range, class2_mle(1), class2_mle
        (2));


    % figure;

    % scatter(x_values, data);
```

```
41
42      % Verify the classes are divided
43      figure;
44
45      hold on;
46
47      yyaxis left;
48      scatter(class1_data, x_values, 'r');
49      scatter(class2_data, x_values, 'b');
50
51      yyaxis right;
52      plot(estimate_range, class1_pdf, 'r');
53      plot(estimate_range, class2_pdf, 'b');
54      legend('Iris Setosa', 'Iris Versicolor');
55      hold off;
56
57      % Plot the likelihood
58      figure;
59      hold on;
60      plot(estimate_range, class1_pdf);
61      plot(estimate_range, class2_pdf);
62      xlim([1, 10]);
63
64      title('Likelihoods');
65      xlabel('x');
66      ylabel('P(x|C_i)');
67
68      p_class1 = class1_pdf ./ (class1_pdf + class2_pdf);
69      p_class2 = class2_pdf ./ (class1_pdf + class2_pdf);
70
71      figure;
72      hold on
73
74      plot(estimate_range, p_class1);
75      plot(estimate_range, p_class2);
76
77      title('Posteriors');
78      xlabel('x');
79      ylabel('P(x|C_i)');
80
81      hold off;
82
83  end
```

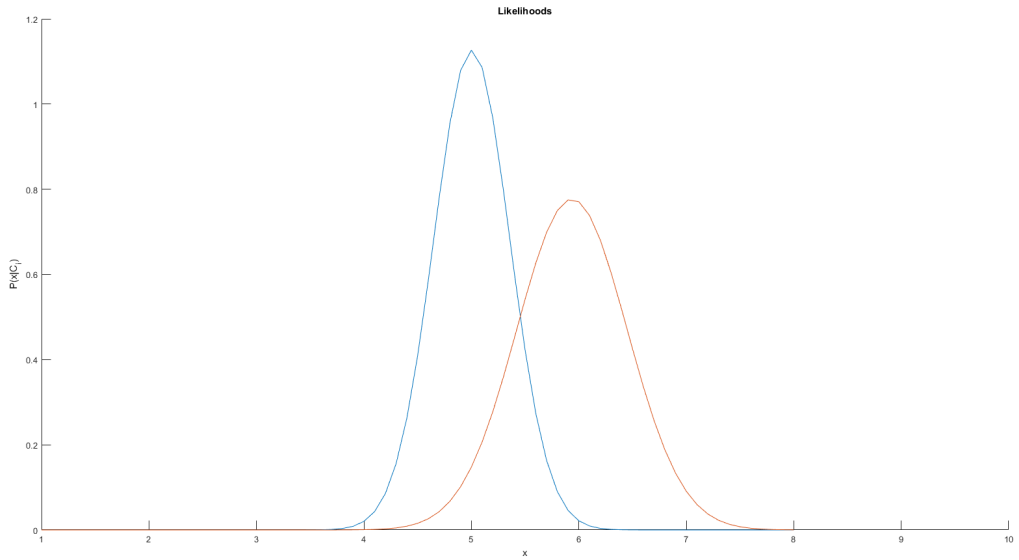

Figure 3: Likelihoods

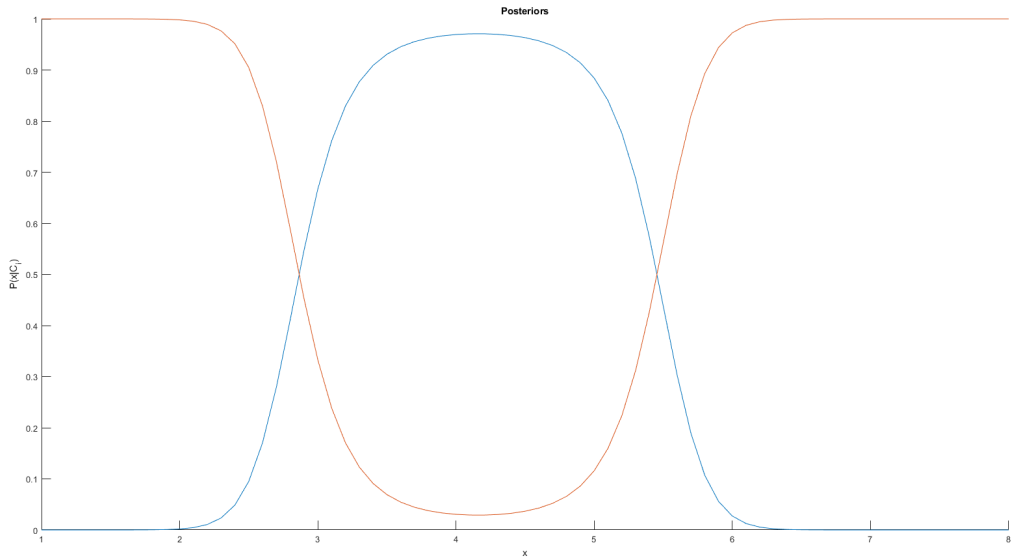

Figure 4: Posteriors

Question 3.1

Question 3.2

Question 3.5

The computed KL values are listed in the table below

M and H1	-1.1571
M and K1	-0.6840
M and K2	1.2875

There was an issue encountered while finding these variables, it was caused by the 0 values in some of the bins of the histogram estimator, this was corrected by

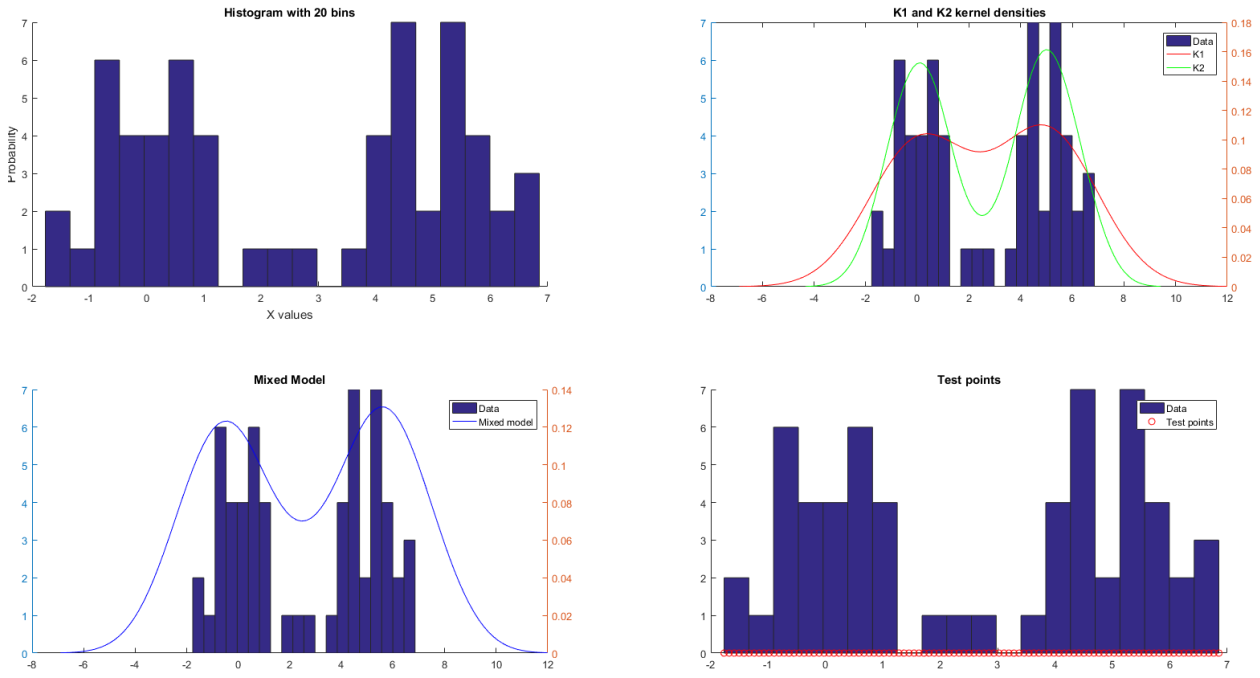

Figure 5: Q5