Rubric Details

Maximum Score	45 points

(i)a ↑
5 possible points (11.11%)

.

3D scatter plot of data. Comment on whether plane or curve. Marks deducted if plot(s) do not axes labels or labels are not legible (e.g. too small to read easily).   **0 – 5**

(i)b ↑
10 possible points (22.22%)

.

Map from two input features to all powers up to 5 using polynomialFeatures(5) or equivalent. Select range of C appropriately. Give parameter values. Provide enough information to explain what parameters/features the reported values correspond to e.g. give model, list of features used and parameter value corresponding to each feature. Comment on how the parameter values change with C - expect to see some zero parameter values and number of these to decrease as C increases.   **0 – 10**

(i)c ↑
5 possible points (11.11%)

.

Generate predictions and plot together with training data. Predictions should span region beyond training data so that generalisation behaviour can be seen - if not then lose marks. Plot should be clear and easy to read - deciding on how to do this is part of the assignment and needs some judgement. Comment on impact of C on predictions - as C increases expect to see more "wobbles" in regions beyond training data. Marks deducted if plot(s) do not axes labels or labels are not legible (e.g. too small to read easily).

0 – 5

(i)d ↑
5 possible points (11.11%)

.

Explain what overfitting and underfitting are - the explanation should be in own words and demonstrate understanding. Discuss choice of C with reference to this. Expect that for small C there will be under-fitting since all model parameters will be zero, and for large C there will be overfitting as parameters associated with higher powers become non-zero.

0 – 5

(i)e ↑
5 possible points (11.11%)

.

As above but for L2 penalty. Again, need to give enough model detail to allow results to be understood (a list iof raw parameter values without explanation is not enough). Expect that will no longer see zero parameter values but rather that magnitude of parameter values will decrease as C is made small - its essential to comment upon this as its the main difference between using L1 and L2 regularisation.

0 – 5

(ii)a ↑
5 possible points (11.11%)

.

Plot of mean and standard deviation of prediction error vs C. Need to choose range of C, explain choice. Marks deducted if plot(s) do not axes labels or labels are not legible (e.g. too small to read easily) and/or range of C values chosen is poor (too small, or so big that can't see what a good value of C is). Its imporant to explain what measure is used for prediction error - the obvious choice is the cost function used for training, but other choices are fine so long as clearly explained. Use of default 'score' in sklearn crossval function, without explanation, means v low mark (its important to demonstrate

0 – 5

understanding of what the numbers being plotted really are).

(ii)b ↑
5 possible points (11.11%)

.

Choose C and explain why. If using a cost (where a lower value is better) then typically want to choose smallest C where the cost first starts to flatten out. If using sklearn score (where higher is better) then this is flipped.

0 – 5

(ii)c ↑
5 possible points (11.11%)

.

As above but for L2 penalty.

0 – 5