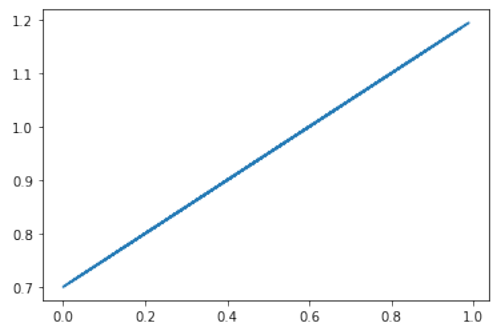
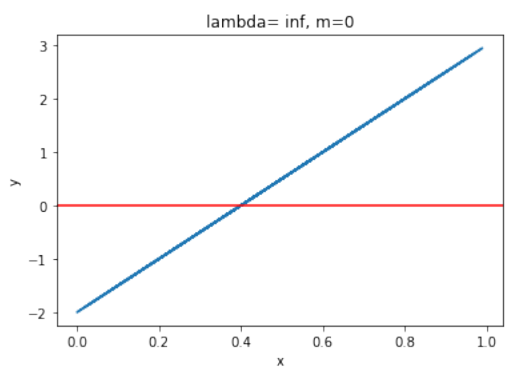
Problem 1. Page 322 Q2

$$\hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(m)}(x)]^2 dx) \tag{1}$$



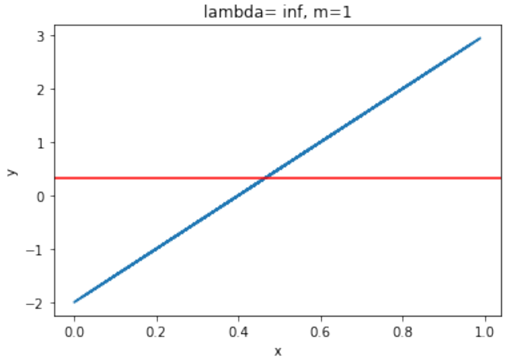
a) $\lambda = \infty$, $m = 0$

- As $\lambda \rightarrow \infty$ the penality term has now become paramount, causing g(x) $\rightarrow$ 0.
  - $\hat{g} = 0$
- for m = 0:
  - $\hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(0)}(x)]^2 dx) \rightarrow \hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g(x)]^2 dx)$



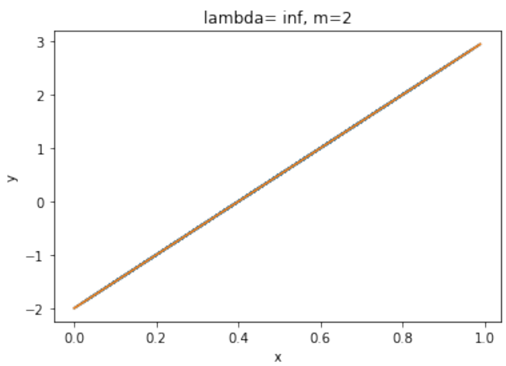
b) $\lambda = \infty$, $m = 1$

- As $\lambda \rightarrow \infty$ the penality term has now become paramount, causing g'(x) $\rightarrow$ 0.
  - $\hat{g} = c = \frac{1}{n}\sum_{i=1}^{n} y_i$
- for m = 1:
  - $\hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(1)}(x)]^2 dx) \rightarrow \hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g'(x)]^2 dx)$



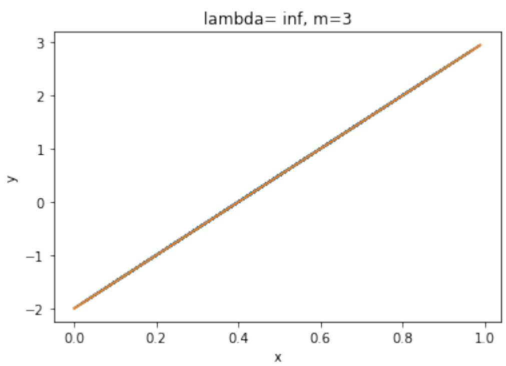
c) $\lambda = \infty, m = 2$

- As $\lambda \to \infty$ the penality term has now become paramount, causing g' '(x) $\to$ 0.
  - $\hat{g} = ax + b$
- for m = 2:
  - $\hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(2)}(x)]^2 dx) \to \hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g''(x)]^2 dx)$



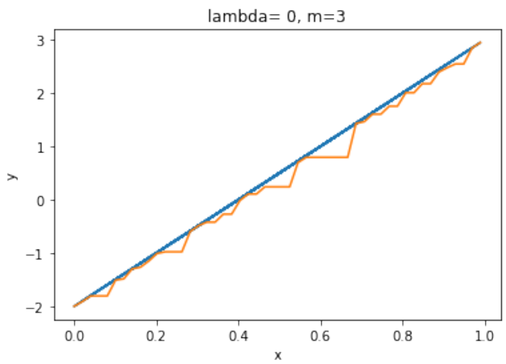
d) $\lambda = \infty, m = 3$

- As $\lambda \to \infty$ the penality term has now become paramount, causing g' ' '(x) $\to$ 0.
  - $\hat{g} = ax^2 + bx + c$
- for m = 3:
  - $\hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(3)}(x)]^2 dx) \to \hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g'''(x)]^2 dx)$



e) $\lambda = 0, m = 3$

- As $\lambda = 0$ the penality term no longer play any role, so g now becomes an interpolating spline.
- for m = 3:
  - $\hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(3)}(x)]^2 dx) \rightarrow \hat{g} = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g'''(x)]^2 dx)$



Problem 1. Page 323 Q5

$$\hat{g}_1 = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(3)}(x)]^2 dx)$$
$$\hat{g}_2 = \underset{g}{argmin}(\sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int [g^{(4)}(x)]^2 dx)$$

a) As $\lambda \rightarrow \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training RSS?

- $\hat{g}_2$ will likely have a smaller training RSS since it has a higher polynomial order, therefore, making the curve more flexible to fit the training data set.

b) As $\lambda \rightarrow \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller test RSS?

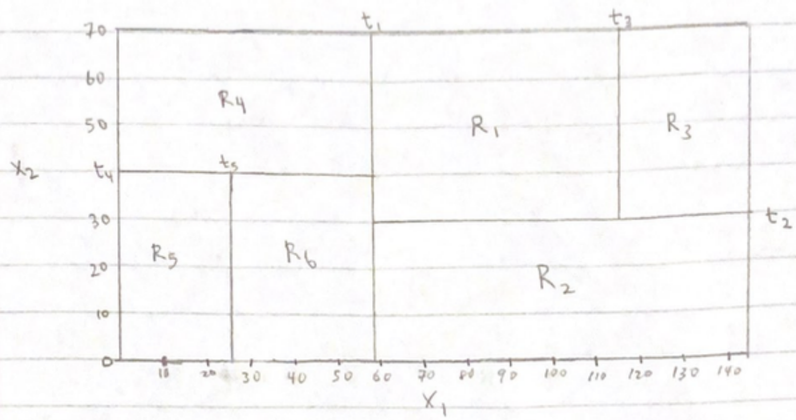
- We cannot give a definite answer to which curve will hae a smaller test RSS, since we do not know the true relationship between the x and y valuables. The $\hat{g}_2$ extra fexiblity could cause it to overfit and have a larger tet RSS, however, if extra complexity is necessary then $\hat{g}_2$ will result in a smaller test RSS. In the otherhand, $\hat{g}_1$ could underfit if the extra complexity is required, resulting in $\hat{g}_1$ having a larger test RSS value.

c) For $\lambda = 0$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training and test RSS?

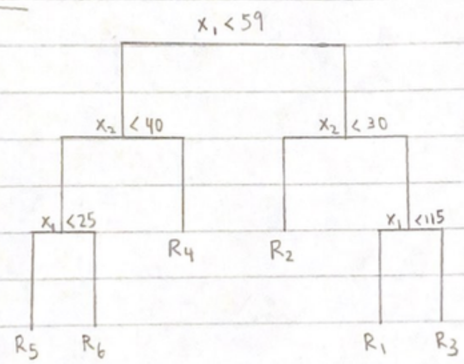
- if $\lambda = 0$ then both $\hat{g}_1$ and $\hat{g}_2$ will be equal to each other, therefore, they will have the same training and test RSS.

Problem 2. Page 361 Q1

2) Page 361 Q1 : example of a partition of 2-D feature space that could result from a recursive binary splitting



Decision tree



Problem 2. Page 362 Q5

(1) First Approach: Majority Vote

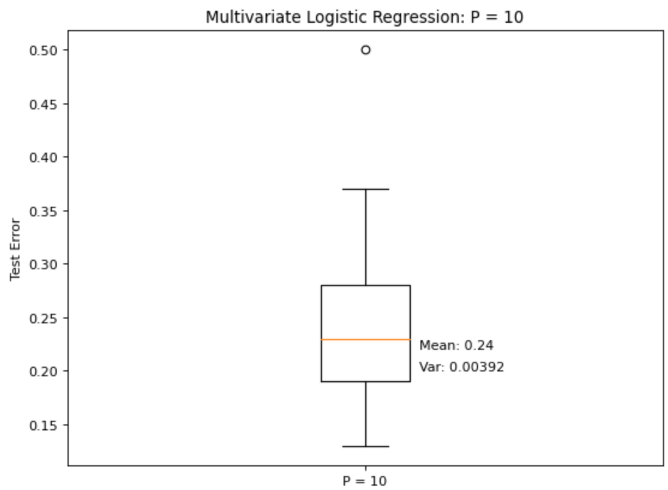
The final classification under the majority vote approach is RED because it is the class that occurs more often in the 10 probability estimates we have (6 are predicted to be red while 4 are predicted to be green).

(2) Second Approach: Average Probability

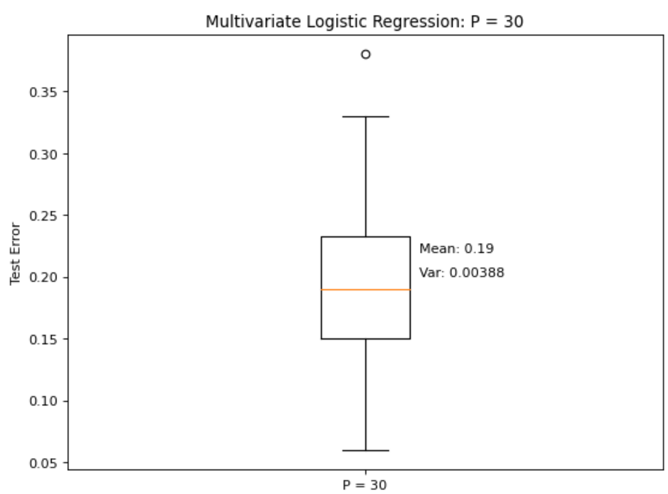
$(0.1 + 0.15 + 0.2 + 0.2 + 0.55 + 0.6 + 0.6 + 0.65 + 0.7 + 0.75)/10 = 0.45$

Since the average of the 10 estimates of P(class is red | X) is 0.45 (< 0.5), the final classification under the average probability approach is GREEN.

Problem 3a.
The mean of the test errors is 0.24
The variance of the test errors is 0.00392
The boxplot for the test errors is shown below.



Problem 3b.
The mean of the test errors is 0.19
The variance of the test errors is 0.00388
The boxplot for the test errors is shown below.



As x_i goes from $x_i \in \mathbb{R}^{10}$ to $x_i \in \mathbb{R}^{30}$ the computation time also increases, the time for P = 10 was (10 minutes) and P = 30 was (50 minutes). In addition, as P increases, the values for the test error become smaller as well. As expected, we observed that the mean and variance for the test error for P = 30 to be smaller than the mean and variance for P = 10. We would expects over-fitting the model to be less of an issue as the sample size becomes larger, thus the values of the test error decreases as P get bigger.

Pledge:
Please sign below (print full name) after checking (✓) the following. If you can not honestly check each of these responses, please email me at kbala@ucdavis.edu to explain your situation.

- We pledge that we are honest students with academic integrity and we have not cheated on this homework. ✓
- These answers are our own work. ✓
- We did not give any other students assistance on this homework. ✓

- We understand that to submit work that is not our own and pretend that it is our is a violation of the UC Davis code of conduct and will be reported to Student Judicial Affairs. ✓
- We understand that suspected misconduct on this homework will be reported to the Office of Student Support and Judicial Affairs and, if established, will result in disciplinary sanctions up through Dismissal from the University and a grade penalty up to a grade of "F" for the course. ✓

Team Member 1: Katherine Shaw Team Member 2: Truc Le