Q2

Part 1

(1) Since m = 0, which means the f term in the second term should be f(x); lambda = infinity, hence, when we are minimizing f1_hat, the second term(which penalizes the curvature of the function) should be enforced to be 0, that is f(x) = 0. It has zero polynomial(or undefined degree of polynomial) since the function f(x) is actually a zero constant.

(2) m = 1 gives us the first derivative f'(x). Similar to (1), lambda = infinity, which means the second term which penalizes curvature of the function should be enforced to zero, that is, f'(x) = 0. f(x) is a constant, hence the degree of polynomial is 0. f1_hat in this case is basically a least square regression.

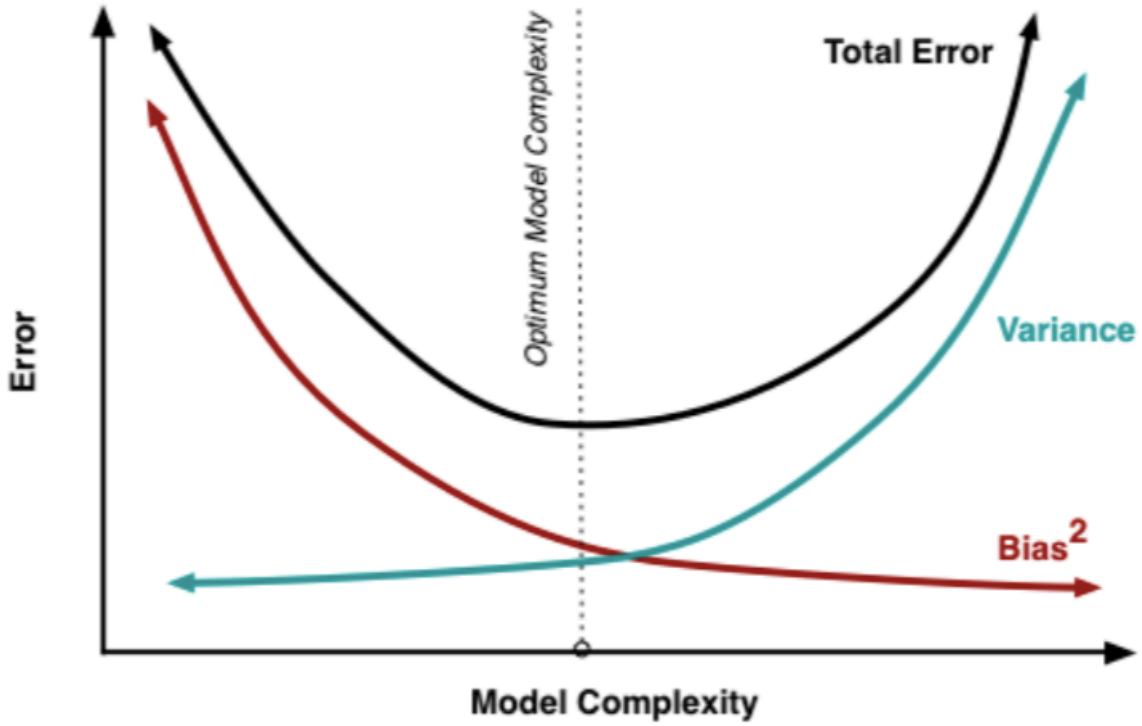
(3) m = 2 gives us the second derivative f''(x). lambda = infinity means when we are minimizing f1_hat, the second term(which penalizes the curvature of the function) should be enforced to be 0. In this case, f''(x) = 0. f(x) is a polynomial of degree 1, takes the form of f(x) = ax + b, where a, b are constant. f1_hat is a linear least square regression in this case.

(4) m=3 gives us the third derivative f'''(x). Lambda = 0 which basically drop the second term which penalizes the curvature of the function. f1_hat only tries to minimize the first term which measures the closeness of the model to the data. f(x) will interpolate all data points which leads to overfitting(a low bias but very high variance). f(x) can have an arbitrarily large polynomial degree.

Part 2

(1) When lambda goes to infinity, both f(m)(x) and f(m+k)(x) will be enforced to 0. Since k is a positive integer, f(m+k)(x) is a higher order degree derivative, which means the corresponding f(x) function has more degree of polynomial. This leads to f2_hat to have a higher variance, lower bias. Hence f2_hat has smaller training RSS.

(2) When lambda goes to infinity, the argument is similar to (1), f1_hat has a lower degree polynimial, hence f1_hat has higher bias, lower variance. f2_hat has higher variance, lower bias. But since we are comparing test RSS, according to the bias-variance tradeoff(the plot taken from internet shown below), the test RSS depends on which model has a smaller bias^2 + variance, which we cannot compare without sufficient conditions here.



(the x axis should be the degree of polynomial in this case)

(3) When lambda = 0, both f1 and f2 interpolate data points as much as they can without the constraints of the curvature. f1_hat and f2_hat will both overfit(a very low bias but very high variance). f1_hat and f2_hat will have the same training RSS, the same test RSS, since the objective functions are basically the same without the second penalization term.

In [ ]: