Polynomial Regression

Consider the following polynomial regression model:

$Y = a + b_1x + b_2x^2 + e$

Want to penalize an over complex function

How do we define the complexity of a regression model?

A: One method is to define complexity as a function of the size of the coefficients.

Ex1: $E|b_i|$ this is called the L1-norm
Ex2: Eb_i^2 this is called the L2-norm also called regularization keeps model as simple as possible L2 is more common. We have to take into account the complexity of the model

These measures of complexity lead to the following regularization techniques:

L1 regularization: $y = E\ B_ix_i + e$ st. $E|B_i| < s$
L2 regularization: $y = E\ B_ix_i + e$ st. $E|B_i^2| < s$

S is the total amount of cost, the error due to having a complex model should be less than the error of our samples

Regularization refers to the method of preventing overfitting by explicitly controlling model complexity.

These regularization problems can also be expressed as:

L1 regularization (Lasso): look up formula
L2 regularization (Ridge): look up formula

This (Lagrangian) formulation reflects the fact that there is a cost associated with regularization.

These measures of complexity lead to the following regularization techniques:

L1 regularization: $y = E\ B_ix_i + e$ st. $E\ |B_i| < s$
L2 regularization: $y = E\ b_ix_i + e$ st. $Eb_i^2 < s$

What are bias and variance?
Bias refers to predictions that are systematically inaccurate
Variance refers to predictions that are generally inaccurate