Training/Testing and Regularization

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

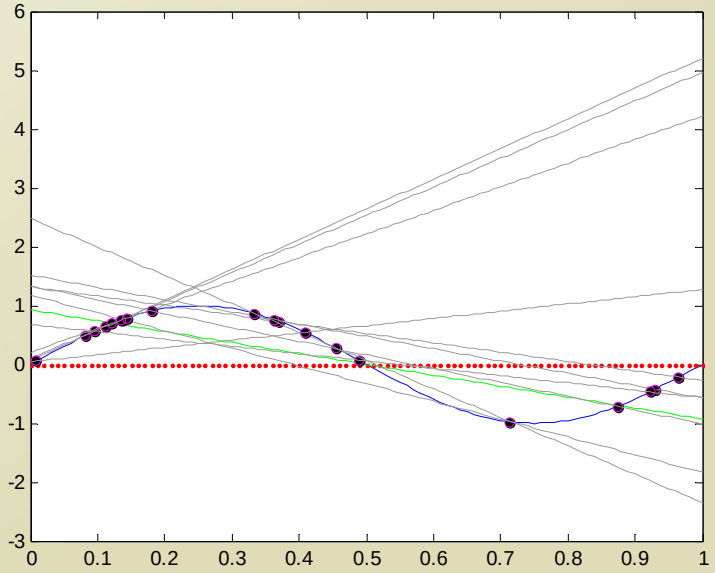
Weekly Objectives

- Understand the concept of bias and variance
  - Know the concept of over-fitting and under-fitting
  - Able to segment two sources, bias and variance, of error
- Understand the bias and variance trade-off
  - Understand the concept of Occam's razor
  - Able to perform cross-validation
  - Know various performance metrics for supervised machine learning
- Understand the concept of regularization
  - Know how to apply regularization to
    - Linear regression
    - Logistic regression
    - Support vector machine

MODEL REGULARIZATION

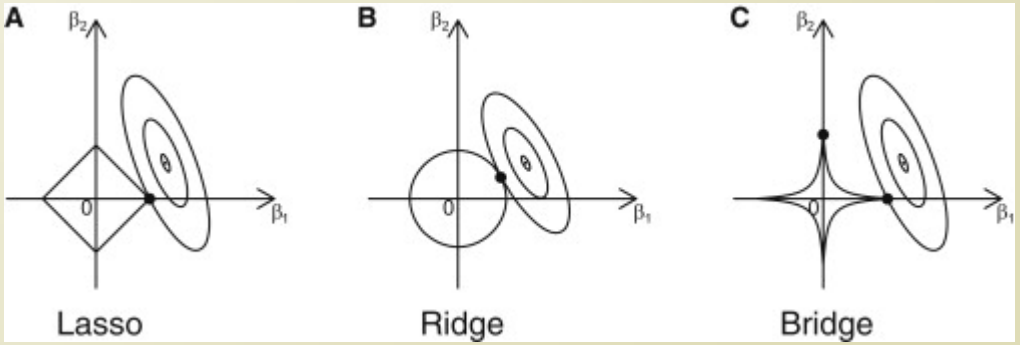
Concept of Regularization

- Disaster in terms of variance
- With regularization
  - We sacrifice the perfect fit
    - Reducing the training accuracy
  - We increase the potential fit in the test
  - Because of the increased model complexity, the bias tends to decrease a little bit
  - Eventually, regularization is another constraint for models
    - Existing constraint?
    - Minimizing error from training set
- We add a new term to the MSE



Formal Definition of Regularization

- Regularization is another constraint for the regression
  - The below J(B) is the regularization function to minimize
  - B is the weight of the regression model except the constant term
- There are diverse regularization
  - L1 Regularization == Lasso regularization
    - The first order
  - L2 Regularization == Ridge regularization
    - The second order
  - Depends on the order of the regularization term
    - The order determines the shape of the loss function



$$E(w) = \frac{1}{2}\sum_{n=0}^{N}(train_n - g(x_n, w))^2 + \frac{\lambda}{2}\|w\|^2$$

$$E(w) = \frac{1}{2}\sum_{n=0}^{N}(train_n - g(x_n, w))^2 + \lambda|w|$$