Machine Learning

Regularization

Regularization

- In the ordinary least squares method, the coefficients β_j are unconstrained. Hence, they can explode and make model susceptible to high variance
- To control variance, we may **regularize** the coefficients to control how large the coefficients grow
- Regularization (also known as shrinkage methods) shrink the coefficients of the attributes and lead us towards simpler yet effective models
- Some coefficients shrink to zero, thus also help reduce curse of dimensionality
- Two regularization methods:
  - Ridge
  - Lasso

Ridge Regression

- Ridge regression addresses some of the problems of Ordinary Least Squares by imposing a penalty on the size of the coefficients

- In linear regression, the following cost function in minimized

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2$$

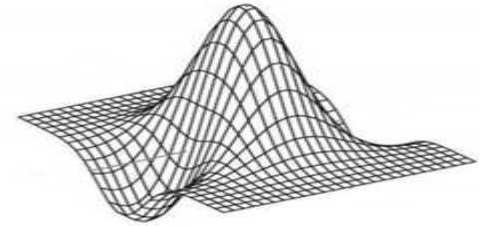
- In Ridge regression, the following cost function in minimized

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2$$

- The term $\lambda\sum_{j=1}^{p}\beta_j^2$ acts like penalty term
  - When Lambda is large, coefficients are suppressed significantly
  - When it is 0, the cost function becomes same as linear regression cost function

Benefits of Regularization

- Unconstrained coefficients along with curse of dimensionality may results in large magnitude coefficients which results in a complex surface / model.



- This complex surface has the data points occupying the peaks and the valleys

- The model gives very high accuracy in training but poor result in testing and the testing scores also vary a lot from one sample to another.

- Such models are overfitted and do not generalize

Benefits of Regularization

- In Ridge Regression, the algorithm while trying to find the best combination of coefficients which minimize the SSE on the training data, is constrained by the penalty term

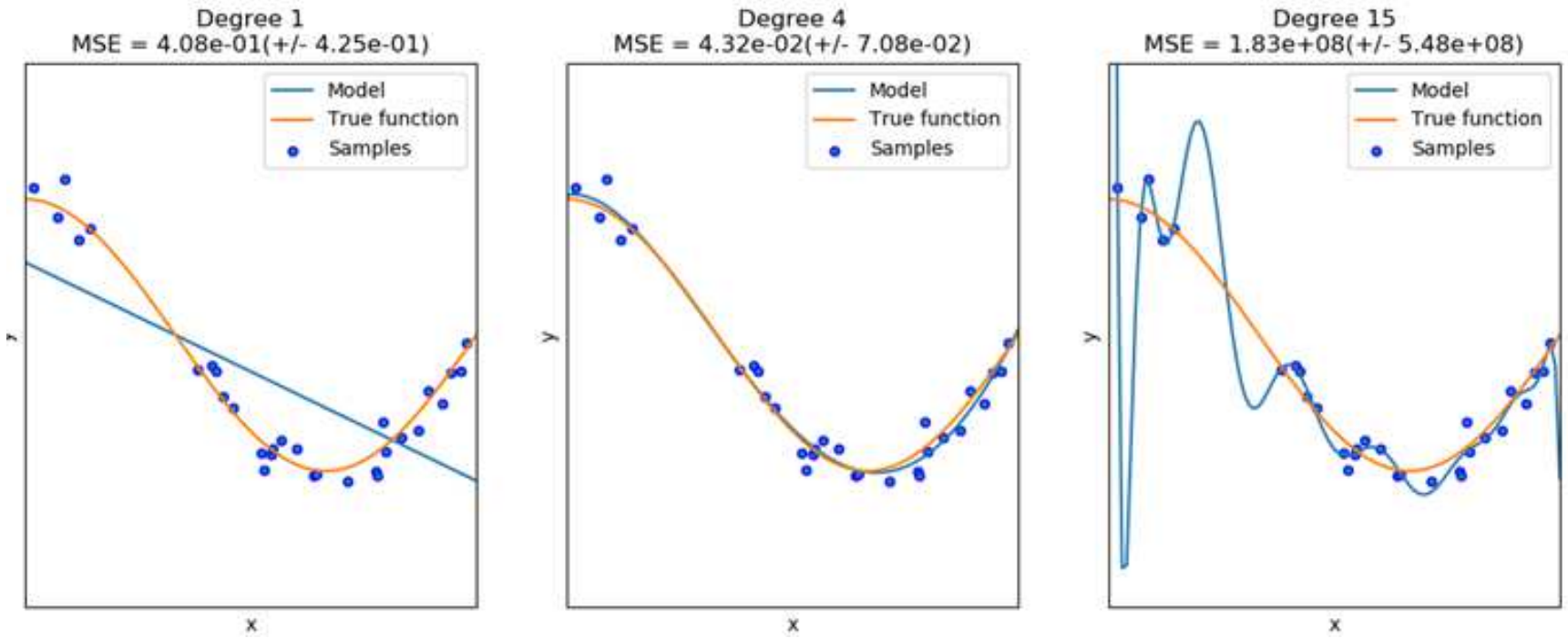
- The penalty term is akin to cost of magnitude of the coefficients. Higher the magnitude, more the cost.

- So to minimize the cost, the coefficient are suppressed

- The resulting surface tends to be relatively smoother. This model will give less accuracy in training as compared to unconstrained model. But such model tend to perform better on test data than unconstrained model.

- Thus the regularized model will generalize better

Which is Right Fit?



Lasso Regression

- Lasso Regression is similar to the Ridge regression with a difference in the penalty term. In Ridge, the penalty has coefficients raised to power 2. In Lasso, the coefficients are raised to power 1. It is also known as L1 norm.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$$

- In Ridge, values of coefficients move towards zero but may not become zero. In Lasso Regression penalty process will make many of the coefficients 0.

- Thus Lasso regression results in dropped dimensions

Exercise

- Ridge and Lasso - Exercise