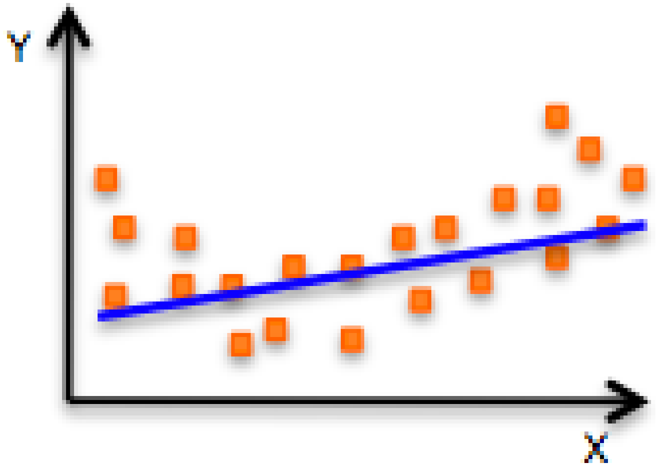
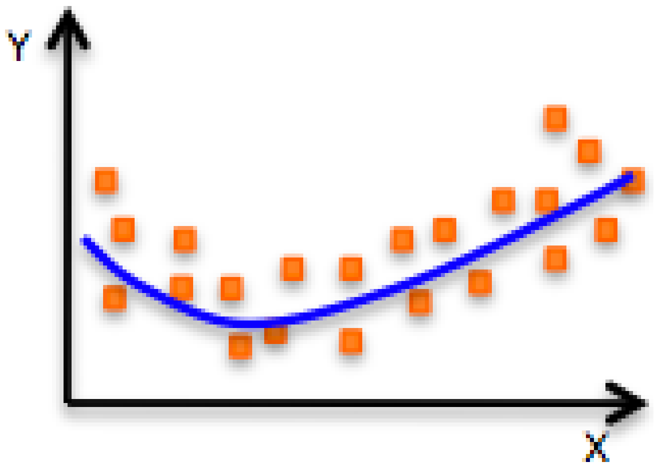
L2 Regularization Derivate

Why Regularization?

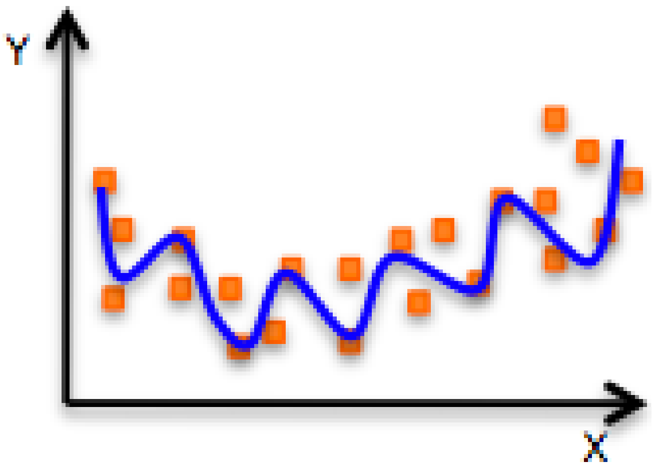
- Constraint the parameter values
- Avoid over-fitting phenomenon



Underfitting



Just right!



overfitting

L2 Regularization in Linear Regression

- Prediction

$$\hat{y} = \boldsymbol{x}^T\boldsymbol{\beta} = \beta_0 + \sum_i x_i\beta_i$$

- Original Objective

$$\min J(\boldsymbol{\beta}) = \min \sum_{(x,y)} \left(\boldsymbol{x}^T\boldsymbol{\beta} - y\right)^2$$

- L2-Regularized Objective

$$\min \sum_{(x,y)} \left(\boldsymbol{x}^T\boldsymbol{\beta} - y\right)^2 + \lambda||\boldsymbol{\beta}||^2$$

Derivative with L2 Regularization

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{(\boldsymbol{x},y)} 2\boldsymbol{x} \cdot \left(\boldsymbol{x}^T \boldsymbol{\beta} - y\right) + \frac{\partial \lambda \|\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}}$$

$$= \sum_{(\boldsymbol{x},y)} 2\boldsymbol{x} \cdot \left(\boldsymbol{x}^T \boldsymbol{\beta} - y\right) + 2\lambda\boldsymbol{\beta}$$

Closed-form with L2-Regularization

$$\nabla_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) = 0$$

$$\nabla_{\boldsymbol{\beta}} \left(\frac{1}{2} ||\boldsymbol{X\beta} - y||^2 + \frac{\lambda}{2} ||\boldsymbol{\beta}||^2 \right) = 0$$

$$\nabla_{\boldsymbol{\beta}} \left(\frac{1}{2} (\boldsymbol{X\beta} - y)^T (\boldsymbol{X\beta} - y) + \frac{\lambda}{2} \boldsymbol{\beta}^T \boldsymbol{\beta} \right) = 0$$

$$\boldsymbol{X}^T \boldsymbol{X\beta} - \boldsymbol{X}^T \boldsymbol{y} + \lambda \boldsymbol{\beta} = 0$$

$$\boldsymbol{X}^T \boldsymbol{X\beta} - \boldsymbol{X}^T \boldsymbol{y} + \lambda \boldsymbol{I\beta} = 0$$

Closed-form with L2-Regularization (Cont'd)

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}^T\boldsymbol{y} + \lambda\boldsymbol{I}\boldsymbol{\beta} = 0$$

$$\boldsymbol{X}^T\boldsymbol{y} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)\boldsymbol{\beta}$$

$$\boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$