CIS 678 - Machine Learning

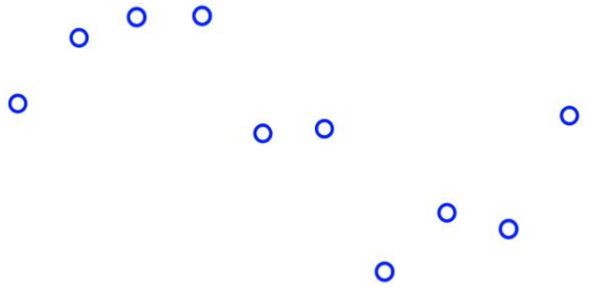
- Model Regularization (part 2)

Plan

- Regularization
  - Theory
  - Practical - Notebook presentation

Non linear data/function

- Does this data points seem familiar matching a known function?

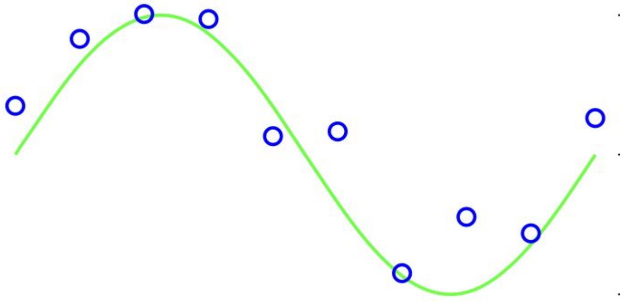


Non linear data/function

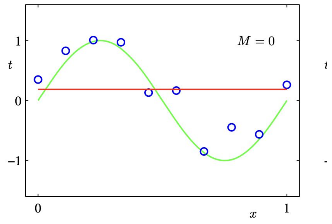
- Does this data points seem familiar matching a known function?

    - A Sinusoidal function

$$y(t) = A\sin(\omega t + \varphi) = A\sin(2\pi f t + \varphi)$$

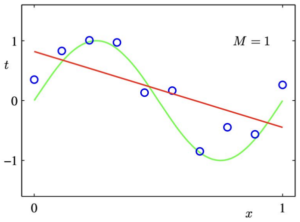


Polynomial Regression with Regularization



$$\hat{y} = \beta_0$$

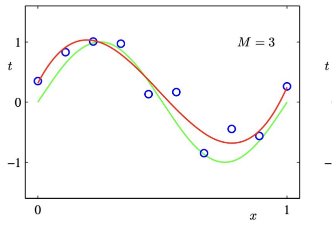
Polynomial Regression with Regularization



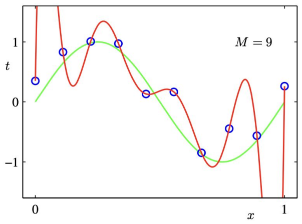
$$\hat{y} = \beta_0 + \beta_1 x$$

Polynomial Regression with Regularization

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$$

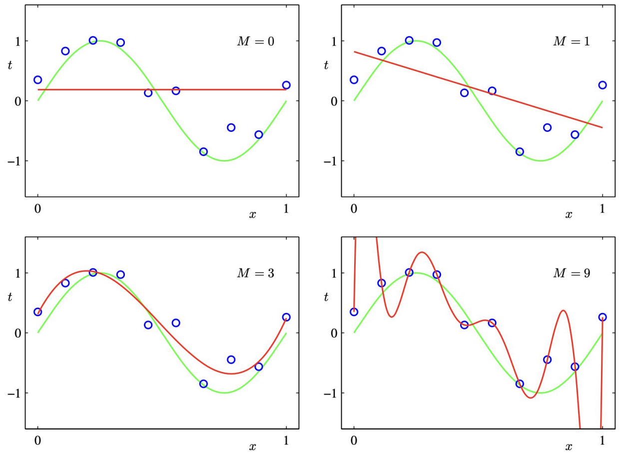


Polynomial Regression with Regularization



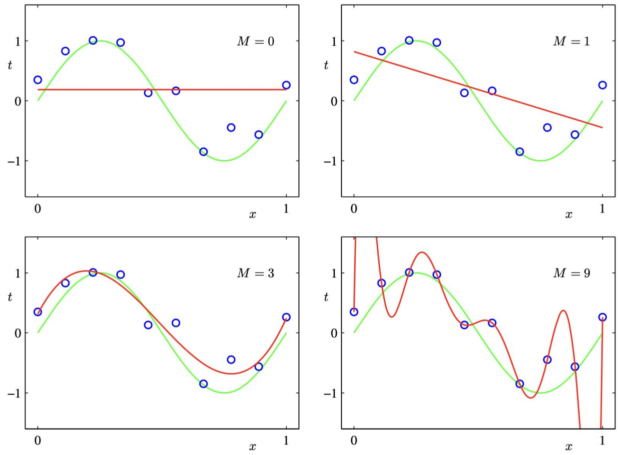
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3 + \ldots$$

Polynomial Regression with Regularization



$$\hat{y} = \beta_0$$
$$\hat{y} = \beta_0 + \beta_1 x$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3 + \ldots$$

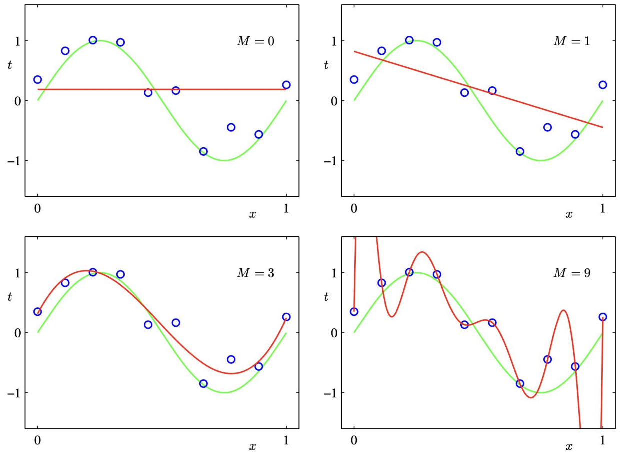
Polynomial Regression with Regularization



$$\hat{y} = \beta_0$$
$$\hat{y} = \beta_0 + \beta_1 x$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3 + \ldots$$

$$\boxed{\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2}$$

Polynomial Regression with Regularization

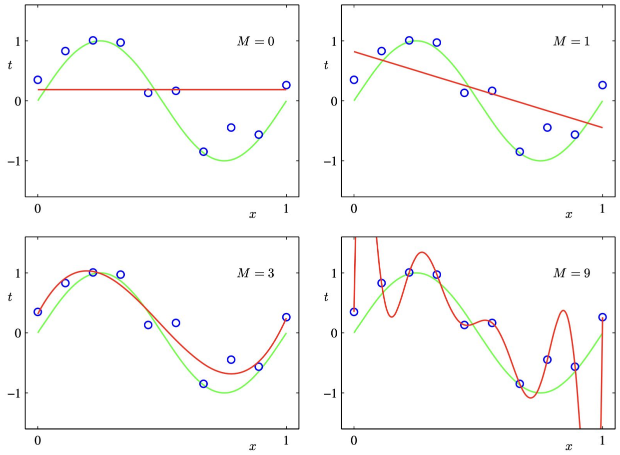


$$\hat{y} = \beta_0$$
$$\hat{y} = \beta_0 + \beta_1 x$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3 + \ldots$$

$$\boxed{\widetilde{E}(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\}^2}$$

decreasing

Polynomial Regression with Regularization



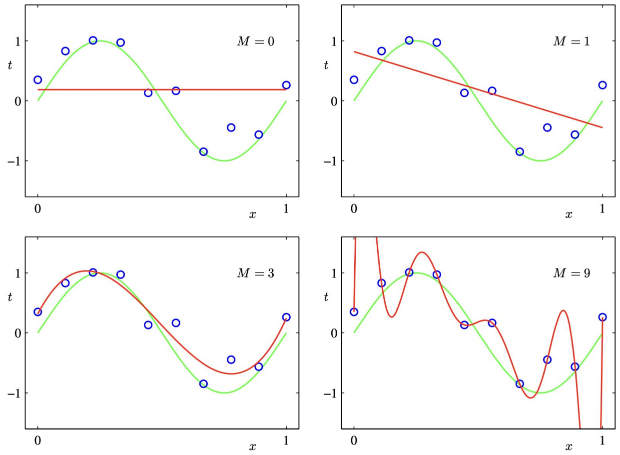
$$\hat{y} = \beta_0$$
$$\hat{y} = \beta_0 + \beta_1 x$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3 + \ldots$$

increasing

$$\boxed{\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{1}{2}\boxed{\|\mathbf{w}\|^2}}$$

decreasing

Polynomial Regression with Regularization



$\hat{y} = \beta_0$

$\hat{y} = \beta_0 + \beta_1 x$

$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$

$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$

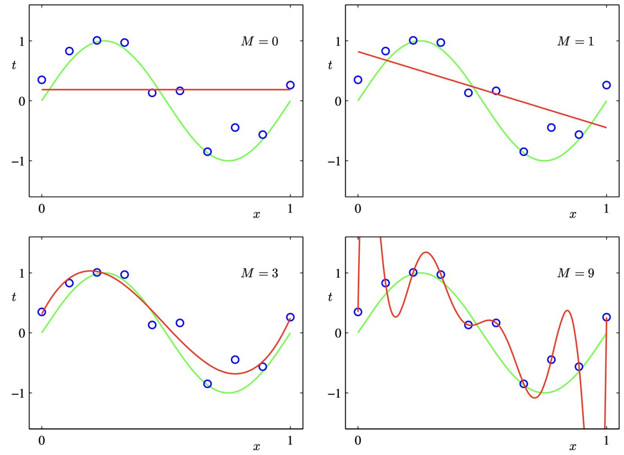
$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots$

Loss function

Regularizer

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{1}{2}\boxed{\|\mathbf{w}\|^2}$$

Polynomial Regression with Regularization



$$\hat{y} = \beta_0$$
$$\hat{y} = \beta_0 + \beta_1 x$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_2 x^3$$
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots$$

Loss function

Regularizer

$$\widetilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}\boxed{\|\mathbf{w}\|^2}$$

Lamda is for balancing these two -
Hyper Parameter of this model

How to find the best HP (Lamda)?

Predictive modeling: Regression (diabetes)