1、Normalization（to find the minimum faster）

$$X_{norm} = \frac{X - \mu}{X_{max} - X_{min}}$$

2、polynomial regression（to draw a decision boundary）

$$\theta^{\mathrm{T}} x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1{}^2 + \theta_4 x_1 \ x_2 \ + \theta_5 x_2{}^2 + \theta_6 x_1{}^3 + \ldots$$

3、sigmoid function & Hypothesis Representation（to predict）

$$\mathrm{g}(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-\theta^{\mathrm{T}} x}}$$

$$h_\theta(x) = g(\theta^T x)$$

4、Cost function（to measure the loss）

$$\mathrm{J}(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

$$\min_\theta J(\theta)$$

5、Gradient Descent & Update parameters（an optimizing）

$$\frac{\partial J(\theta)}{\partial \theta_{\mathrm{j}}} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

$$\theta_{\mathrm{j}} = \theta_{\mathrm{j}} - \alpha\frac{\partial J(\theta)}{\partial \theta_j} \quad \text{for j>=1}$$

6、Regularization(to prevent overfitting)

Cost function

$$\mathrm{J}(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)}\log h_\theta(x^{(i)}) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))] + \frac{\lambda}{2\mathrm{m}}\sum_{j=1}^{n}\theta_j^2$$

Gradient Descent

$$\frac{\partial J(\theta)}{\partial \theta_{\mathrm{j}}} = (\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\ ) + \frac{\lambda}{\mathrm{m}}\theta_j \qquad \text{for j>=1}$$