

Logistic Regression

- Hypothesis Representation

Want

$$0 \leq h_{\theta}(x) \leq 1$$

Sigmoid function:

$$g(z) = \frac{1}{1 + e^{-z}}$$

Logistic function:

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

Interpretation:

$$h_{\theta}(x) = \text{estimated probability that } y = 1 \text{ on input } x$$

- Decision Boundary

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

$$g(z) = \frac{1}{1 + e^{-z}}$$

Suppose

$$\text{predict } y = 1, \text{ if } h_{\theta}(x) \geq 0.5$$

$$\text{predict } y = 0, \text{ if } h_{\theta}(x) < 0.5$$

Decision boundary

$$\theta^T x = 0$$

- Non-linear decision boundaries

$$h_{\theta}(x) = g(-1 + x_1^2 + x_2^2)$$

$$\text{Predict } y = 1, \text{ if } -1 + x_1^2 + x_2^2 \geq 0$$

$$\text{Predict } y = 0, \text{ if } -1 + x_1^2 + x_2^2 < 0$$

Decision boundary

$$x_1^2 + x_2^2 = 1$$

- Cost function

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y^{(i)})$$

$$\text{Cost}(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)), & \text{if } y = 1 \\ -\log(1 - h_{\theta}(x)), & \text{if } y = 0 \end{cases}$$

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

To filter parameters θ :

$$\min J(\theta)$$

To make a prediction given new x :

$$\text{Output } h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

- Gradient descent

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

Want $\min_{\theta} J(\theta)$:

Repeat {

$$\begin{aligned}\theta_j &:= \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \\ &= \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}\end{aligned}$$

}

- One-vs-all

Train a logistic regression classifier $h_{\theta}^{(i)}(x)$ for each class i to predict the probability that $y = i$.

On a new input x , to make a prediction, pick the class i that maximizes

$$\max h_{\theta}^{(i)}(x)$$