ML Week

0x03 Logistic Regression

Jeff Abrahamson

2–5 novembre 2015

Linear regression

- Continuous output
- Normal residues

Logistic regression

- Binary output
- Classification

Logistic regression

- Have: continuous and discrete inputs
- Want: class (0 or 1)

Probabilistic inspiration

$$h_\theta(x) = .75 \iff \text{event has 70\% of being true}$$

Probabilistic inspiration

$$h_\theta(x) = \Pr(y = 1 \mid x; \theta) = 0.70$$

Probabilistic inspiration

So this must be true:

$$\Pr(y = 0 \mid x; \theta) + \Pr(y = 1 \mid x; \theta) = 1$$

Probabilistic inspiration

$$\text{Set } y = 1 \iff h_\theta(x) = \Pr(y = 1 \mid x; \theta)$$

Probabilistic inspiration

Math review:

- $z = (\theta^T x)$
- $\theta^T x \geqslant 0 \iff h_\theta \geqslant 0.5$
- $\theta^T x \geqslant 0 \iff \text{predict } y = 1$

Logistic (sigmoid) function

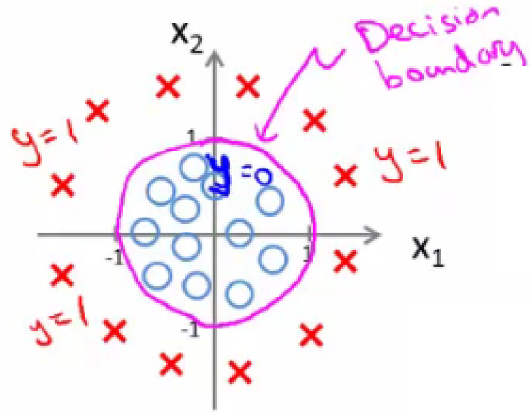
$$g(z) = \frac{1}{1 + e^{-z}}$$

Logistic (sigmoid) function

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

Exercise: plot this

Non-linear decision boundaries



$$h_\theta(x)) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2$$

Andrew Ng

Non-linear decision boundaries

OvA, OvR

OvO

Cost function in logistic regression

In linear regression, we had

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Cost function in logistic regression

In linear regression, we had

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x) - y)^2$$

Cost function in logistic regression

In linear regression, we had

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m} \text{Cost}(h_\theta(x), y)$$

Cost function in logistic regression

Here's a convex cost function:

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

Cost function in logistic regression

Here's a convex cost function:

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

Exercise: Plot this (cost vs y).

Cost function in logistic regression

Here's a convex cost function:

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

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \text{Cost}(h_\theta(x), y)$$

Cost function in logistic regression

Here's a convex cost function:

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

$$J(\theta) = y \cdot \log(h_\theta(x)) + (1 - y) \cdot \log(1 - h_\theta(x))$$

Gradient descent

$$\theta_j \leftarrow \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) \cdot x_j^{(i)}$$

$$\text{for } j = 1, \cdots, n$$

Questions?

`purple.com/talk-feedback`