Machine Learning

Logistic Regression

Jeff Abrahamson

9, 16 décembre 2016 13 janvier, 9 mars 2017

Linear regression

- Continuous output
- Normal residues
- Predict $\hat{y}$ for x given $\{(x_i, y_i)\}$

Logistic regression

- Binary output
- Classification

Logistic regression

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

Probabilistic inspiration

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

Probabilistic inspiration

$$h_\theta(x) = \Pr(y = 1 \mid x; \theta) = 0.75$$

Probabilistic inspiration

So this must be true:

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

Probabilistic inspiration

$$\text{Set } y = 1 \iff h_\theta(x) = \Pr(y = 1 \mid x; \theta) > \frac{1}{2}$$

Probabilistic inspiration

Math review:

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

Logistic (sigmoid, logit) function

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

Logistic (sigmoid, logit) function

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

Exercise: plot this

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

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

Cost function in logistic regression

Here's a convex cost function:

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

Cost function in logistic regression

Here's a convex cost function:

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

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:

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

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

Gradient descent

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

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

null hypothesis

true positive, true negative

false positive, false negative

type I error

(incorrect rejection of null hypothesis)

type II error

(failure to reject null hypothesis)

sensitivity

100% sensitivity = no false negatives

specificity

100% specificity = no false positives

Precision

$$P = \frac{TP}{TP + FP}$$

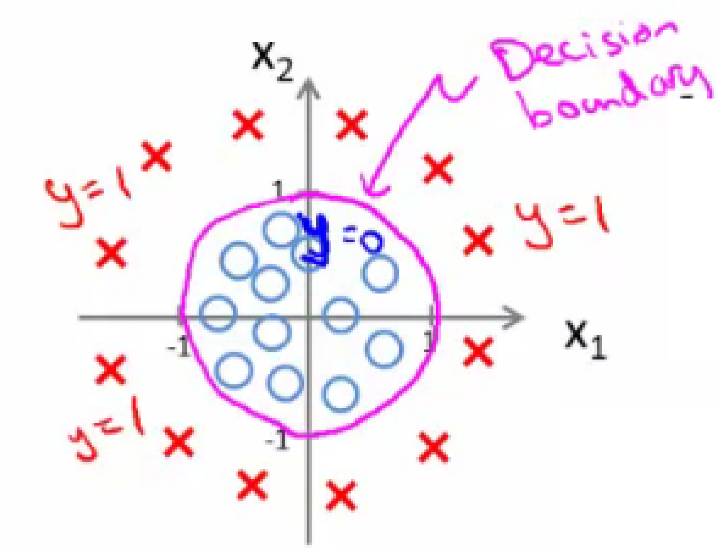
Recall

$$R = \frac{TP}{TP + FN}$$

F1 score

$$F1 = \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Non-linear decision boundaries



Non-linear decision boundaries

OvA = OvR

OvO

Non-linear decision boundaries

One vs All = One vs Rest

One vs One

questions?