Machine Learning Week-5

ramay7

March 13, 2017

Contents

1 Hypothesis

Logistic regression:

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

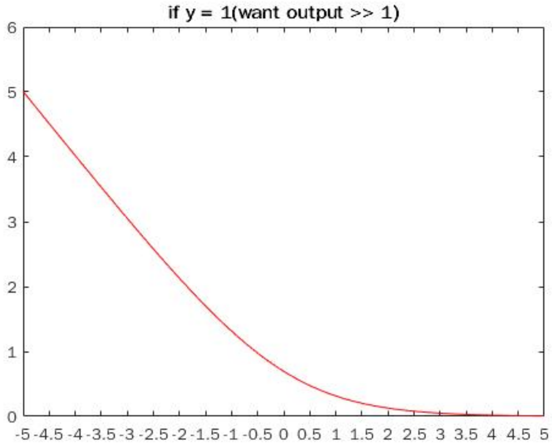
support vector machine hypothesis:

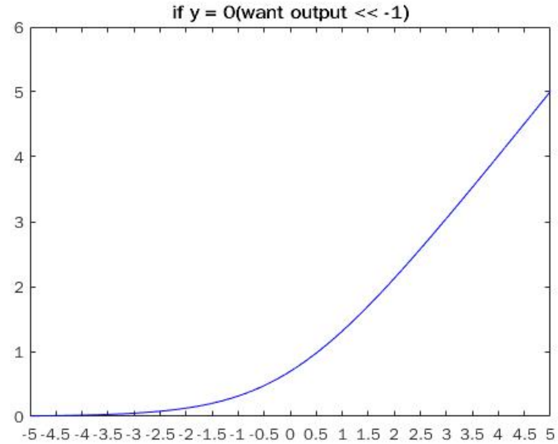
$$\min_{\theta} C\left[\sum_{i=1}^{m} y^{(i)}cost_1(\theta^T x^{(i)}) + (1-y^{(i)})cost_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

for some obvious reason, we can know: $C = \frac{1}{\lambda}$

2 SVM Decision Boundary

Then the next two picturs show the curve of logistic function while in the case of y = 1 and y = 0:





So we can simplify the hypothesis as:

$$\min_{\theta} \frac{1}{2} \sum_{j=1}^{n} \theta_j^2$$

Due to the fact that $\theta^T x^{(i)} \geq 1$ if $y^{(i)} = 1$ and $\theta^T x^{(i)} \leq -1$ if $y^{(i)} = 0$.

3 Kernels and Similarity

$$f_1 = similarity(x, l^{(1)}) = \exp(-\frac{||x - l^{(1)}||^2}{2\delta^2})$$

If $x \approx l^{(1)}$, $f_1 = 1$; If x is far from $l^{(1)}$, $f_1 = 0$.

4 SVM Parameters

$C(= \frac{1}{\lambda})$.

- Large C: Lower bias, high variance
- Small C: Higher bias, low variance

δ^2:

- Large δ^2: Feature f_i vary more smoothly; Higher bias, lower variance
- Small δ^2: Feature f_i vary less smoothly; Lower bias, higher variance

5 Logistic Regression vs SVMs

n = number of features, m = number of training examples If n is large(relative to m):

- use logistic regression, or SVM without a kernel(”linear kernel”)

If n is small, m is intermediate:

- use SVM with Gaussian Kernel

If n is small, m is large:

- Create/add more features, then use logistic regression or SVM withut a kernel

Neural network is likely to work well for most of these settings, but may be slower to train.