

Coursera Machine Learning Week7

MachineLearning Coursera

Coursera Machine Learning Week7

支持向量机SVM(Support Vector Machine)

- 1.从逻辑回归到SVM
- 2.大间距分类器 Large Margin Intuition
 - 决策边界 Decision Boundary
 - 异常点
- 3.核函数 Kernels
 - 新的特征变量
 - 如何获得核 l^i
 - 核函数的本质
 - 不同的核函数
- 4.选择参数
- 5.多类分类

支持向量机SVM(Support Vector Machine)

1.从逻辑回归到SVM

在Logistics regression中，我们约定的cost function是:

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

这里 $h(x)$ 就是logistics函数，我们用两个近似函数 $cost_1$ 和 $cost_0$ 来替代他们，那么在SVM中
SVM hypothesis :

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

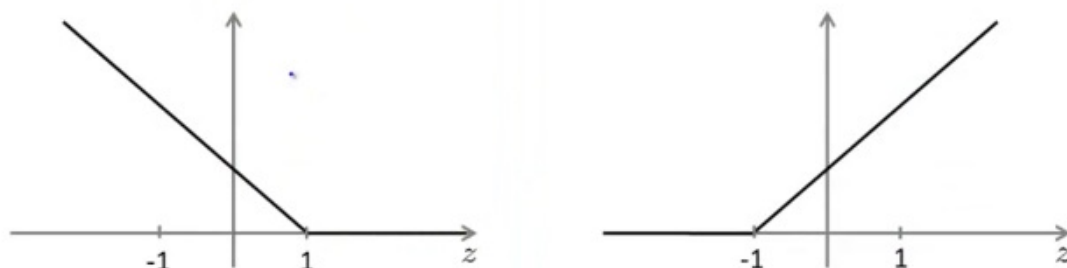
如果相对于样本个数来说，样本的特征个数较大也就是n较大时，使用逻辑回归。

如果n小，m中等，SVM(高斯)。

如果n小，m大，逻辑回归。

Support Vector Machine

$$\rightarrow \min_{\theta} C \sum_{i=1}^m \left[y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$



If $y = 1$, we want $\theta^T x \geq 1$ (not just ≥ 0)

If $y = 0$, we want $\theta^T x \leq -1$ (not just < 0)

2.大间距分类器 Large Margin Intuition

决策边界 Decision Boundary

margin指的是边界离样本的最近的距离，使margin最大化就是使cost function最小化的过程。因为使J变小也就是使 $\frac{1}{2} \|\theta\|^2$ 变小（因为当C较大时我们会使前面那项尽可能为0，由此忽略）。但是因为 $y > 1$ 时要有 $p^i * \|\theta\| > 1$ ，所以要使 θ 小的话， p^i 就得大，也就是样本点离边界的距离要大，也就是margin要大。

异常点

SVM在处理异常点上会根据C的值来给定边界。

3.核函数 Kernels

新的特征变量

限定原有的特征变量组成的向量 $X = (x_i \sim x_j)$ ，选择 l^k ，那么可以使用x和l的相似度 $f^k = \text{similarity}(x, l^k) = \exp(1 \frac{\|x - l^k\|^2}{2\sigma^2})$ 作为新的特征变量。这里的相似函数similarity也被称为核函数Kernel(这里用的是高斯核函数Gaussian Kernel)。

如何获得核 l^i

选择m个样本，将每一个都可以作为一个核。

核函数的本质

就是将原来的特征向量转换为一个新的更适合SVM处理的特征向量。

不同的核函数

- 线性核函数 (没有核函数) $f^i = x^i$
- 高斯核函数 $f_i = \exp(1 \frac{\|x-l^i\|^2}{2\sigma^2})$, where $l^i = x^i$
- 核函数需要满足 Mercer's Throrem从而使用各种数值优化算法。现在其他的核函数已经不多用了，它们有
 1. Polynomial kernel
 2. String kernel
 3. chi-square kernel等等

4.选择参数

$$C = \frac{1}{\lambda}$$

Large C:Lower bias,high variance.

Small C:Higher bias,low variance.

$$\sigma^2$$

Large:f分布更平滑，Higher bias,lower variance.

Small:较陡峭,Lower bias,Higher variance.

5.多类分类

训练K个SVM。