机器学习笔记_阳畅_20160321

一、线性回归

1、假设函数

$h_\theta(x) = \theta_0 + \theta_1 x$

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$

2、代价函数

$J(\theta_0, \theta_1) = \frac{1}{2m} \sum\limits_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$

$$J(\theta_0, \theta_1, \ldots, \theta_n) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

$J(\theta) = \frac{1}{2m} \left[\sum\limits_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum\limits_{j=1}^{n} \theta_j^2\right]$（防止过拟合）

3、梯度下降

$$\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \\
&\qquad (\text{for } j = 1 \text{ and } j = 0) \\
&\}
\end{aligned}$$

$$\begin{aligned}
&\text{repeat until convergence } \{ \\
&\quad \theta_0 := \theta_0 - \alpha \tfrac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \\
&\quad \theta_1 := \theta_1 - \alpha \tfrac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right) \cdot x^{(i)} \\
&\}
\end{aligned}$$

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

$\theta_j := \theta_j (1 - \alpha \frac{\lambda}{m}) - \alpha \frac{1}{m} \sum\limits_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$（防止过拟合）

4、特征缩放（加速迭代）

$-1 \le x_i \le 1$

$x_1 = \frac{size - 1000}{2000}$ $\quad -0.5 \le x_1 \le 0.5, -0.5 \le x_2 \le 0.5$（归一化）

5、选择α

$\ldots, 0.001, 0.000, 0.01, 0.000, 0.1, 0.000, 1, \ldots$

6、转换为线性回归模型

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$
$= \theta_0 + \theta_1 (size) + \theta_2 (size)^2 + \theta_3 (size)^3$

7、最小二乘法

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \\ 1 & 3000 & 4 & 1 & 38 \end{bmatrix}$$

$\theta = (X^T X)^{-1} X^T y$

$$\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \right)^{-1} X^T y$$ （防止过拟合）

二、逻辑回归

1、假设函数

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

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

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

$\theta^T x_{>=0} \quad h_\theta(x) \ge 0.5 \quad y = 1$

$\theta^T x_{<0} \quad h_\theta(x) < 0.5 \quad y = 0$

2、代价函数

$$J(\theta)=\frac{1}{m}\sum_{i=1}^{m}\mathrm{Cost}(h_\theta(x^{(i)}),y^{(i)})$$
$$=-\frac{1}{m}[\sum_{i=1}^{m}y^{(i)}\log h_\theta(x^{(i)})+(1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$
$$J(\theta)=\left[-\frac{1}{m}\sum_{i=1}^{m}y^{(i)}\log(h_\theta(x^{(i)})+(1-y^{(i)})\log 1-h_\theta(x^{(i)})\right]+\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$ （防止过拟合）

3、梯度下降

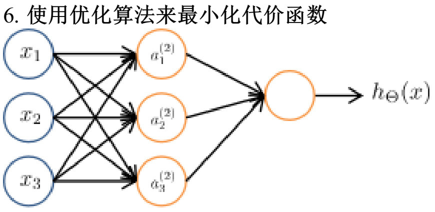
$$\theta_j:=\theta_j-\alpha\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})x_j^{(i)}$$
$$h_\theta(x)=\frac{1}{1+e^{-\theta^Tx}}$$
$$\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})x_0^{(i)}$$
$$\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})x_1^{(i)}-\frac{\lambda}{m}\theta_1$$ （防止过拟合）

4、多分类

$$h_\theta^{(i)}(x)=P(y=i|x;\theta)$$
$$\max_i h_\theta^{(i)}(x)$$

三、神经网络

1. 参数的随机初始化
2. 利用正向传播方法计算所有的 hθ(x)
3. 编写计算代价函数 J 的代码
4. 利用反向传播方法计算所有偏导数
5. 利用数值检验方法检验这些偏导数
6. 使用优化算法来最小化代价函数



假设函数

$$h_\Theta(x)=a_1^{(3)}=g(\Theta_{10}^{(2)}a_0^{(2)}+\Theta_{11}^{(2)}a_1^{(2)}+\Theta_{12}^{(2)}a_2^{(2)}+\Theta_{13}^{(2)}a_3^{(2)})$$

代价函数

$$h_\Theta(x)\in\mathbb{R}^K\quad(h_\Theta(x))_i=i^{th}\text{ output}$$
$$J(\Theta)=-\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K}y_k^{(i)}\log(h_\Theta(x^{(i)}))_k+(1-y_k^{(i)})\log(1-(h_\Theta(x^{(i)}))_k)\right]$$
$$+\frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(\Theta_{ji}^{(l)})^2$$

计算流程

$$\delta^{(3)}=(\Theta^{(3)})^T\delta^{(4)}.*g'(z^{(3)})$$

$g'(z^{(3)})=a^{(3)}.*(1-a^{(3)})$

For $i=1$ to m

Set $a^{(1)}=x^{(i)}$

Perform forward propagation to compute $a^{(l)}$ for $l=2,3,\ldots,L$

Using $y^{(i)}$, compute $\delta^{(L)}=a^{(L)}-y^{(i)}$

Compute $\delta^{(L-1)},\delta^{(L-2)},\ldots,\delta^{(2)}$

$\triangle_{ij}^{(l)}:=\triangle_{ij}^{(l)}+a_j^{(l)}\delta_i^{(l+1)}$

$$D_{ij}^{(l)}:=\frac{1}{m}\triangle_{ij}^{(l)}+\lambda\Theta_{ij}^{(l)}\ \text{if } j\neq 0$$
$$D_{ij}^{(l)}:=\frac{1}{m}\triangle_{ij}^{(l)}\qquad\text{if } j=0$$

$$\frac{\partial}{\partial\Theta_{ij}^{(l)}}J(\Theta)=D_{ij}^{(l)}$$

梯度检查

四、支持向量机

代价函数和假设函数

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

$$\min\frac{1}{2}\sum_{j=1}^{n}\theta_j^2\quad s.t\begin{cases}\theta^T x^{(i)}\geq 1\ if\ y^{(i)}=1\\ \theta^T x^{(i)}\leq -1\ if\ y^{(i)}=0\end{cases}$$

$$f_1=similarity(x,l^{(1)})=e\left(-\frac{\left\|x-l^{(1)}\right\|^2}{2\sigma^2}\right)$$

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

回顾 C=1/λ，因此：

C 较大时，相当于λ较小，可能会导致过拟合，高方差。

C 较小时，相当于λ较大，可能会导致低拟合，高偏差。

σ较大时，导致高方差

σ较大时，导致高偏差

五、k-means

用μ1,μ2,...,μk 来表示聚类中心，用 c(1),c(2),...,c(m)来存储与第 i 个实例数据最近的聚类中心的索引，K-均值算法的伪代码如下：

Repeat {

for i = 1 to m

c(i) := index (form 1 to K) of cluster centroid closest to x(i)

for k = 1 to K

μk := average (mean) of points assigned to cluster k

}