CS229_3

CS229

正规方程

上一小节中，我们使用批量梯度下降算法，通过不断迭代以求得最佳参数θ的值。本小节将介绍另一种方法——正规方程（The Normal Euqations）来计算出最佳参数θ的值。

在介绍正规方程法之前，我们先看看一些基本概念。

Matrix Derivatives

对于一个$m*n$的矩阵到实数的函数映射$f: R^{m*n} \mapsto R$，其关于A的导数为：

$$\nabla_A f(A) = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \cdots & \frac{\partial f}{\partial A_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{m1}} & \cdots & \frac{\partial f}{\partial A_{mn}} \end{bmatrix}$$

其中A为$m*n$的矩阵。

便于理解，我们不妨假设矩阵A为：

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

函数映射$f: R^{2*2} \mapsto R$为：

$$f(A) = \frac{3}{2} A_{11} + 5A_{12}^2 + A_{21}A_{22}$$

根据上述公式，我们可得：

$$\nabla_A f(A) = \begin{bmatrix} \frac{3}{2} & 10A_{12} \\ A_{22} & A_{21} \end{bmatrix}$$

对于$n*n$矩阵A，我们将矩阵A对角线上元素的和定义为矩阵A的迹：

$$trA = \sum_{i=1}^{n} A_{ii}$$

其中若矩阵A为$1*1$，即为一实数，则其迹为本身，$trA=A$。

一些常用性质如下：

$$
\begin{gathered}
trAB=trBA \\
trABC=trCAB=trBCA \\
trA=trA^T \\
tr(A+B)=tr(A)+tr(B) \\
traA=atrA
\end{gathered}
$$

结合矩阵导数的概念有如下性质：

$$
\begin{align}
\nabla_A trAB &= B^T \tag{1} \\
\nabla_{A^T} f(A) &= (\nabla_A f(A))^T \tag{2} \\
\nabla_A trABA^TC &= CAB+C^TAB^T \tag{3} \\
\nabla_A |A| &= |A|(A^{-1})^T \tag{4}
\end{align}
$$

其中等式（1）要求AB为方阵；等式（3）要求ABA^TC为方阵；等式（4）要求矩阵A为非奇异矩阵，即可逆；$|A|$表示矩阵A的行列式。

Least Squares Revisited

好了，现在让我们开始介绍正规方程法，以找到最佳参数θ的值最小化代价函数$J(\theta)$。

在给定训练集中，我们可构建一个维度为$m*n$的矩阵X，其中m为样本个数，n为每个样本的特征变量个数。

$$
X=\begin{bmatrix} -(x^{(1)})^T- \\ -(x^{(2)})^T- \\ \vdots \\ -(x^{(m)})^T- \end{bmatrix}
$$

同样，向量Y为：

$$
Y=\begin{bmatrix} (y^{(1)})^T \\ (y^{(2)})^T \\ \vdots \\ (y^{(m)})^T \end{bmatrix}
$$

根据$h_\theta(x^{(i)})=(x^{(i)})^T\theta$，我们可得：

$$
\begin{aligned}
X\theta - Y &= \begin{bmatrix} (x^{(1)})^T\theta \\ (x^{(2)})^T\theta \\ \vdots \\ (x^{(m)})^T\theta \end{bmatrix} - \begin{bmatrix} (y^{(1)})^T \\ (y^{(2)})^T \\ \vdots \\ (y^{(m)})^T \end{bmatrix} \\
&= \begin{bmatrix} h_\theta(x^{(1)}) - y^{(1)} \\ h_\theta(x^{(2)}) - y^{(2)} \\ \vdots \\ h_\theta(x^{(m)}) - y^{(m)} \end{bmatrix}
\end{aligned}
$$

又因为对于向量z，有$z^Tz = \sum_i z_i^2$。故我们可得：

$$
J(\theta) = \frac{1}{2}(X\theta - Y)^T(X\theta - Y) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x_{(i)}) - y^{(i)})^2
$$

所以，我们对代价函数$J(\theta)$求偏导，可得：

$$
\begin{aligned}
\nabla_\theta J(\theta) &= \frac{1}{2}\nabla_\theta(\theta^TX^T - Y^T)(X\theta - Y) && (1) \\
&= \frac{1}{2}\nabla_\theta(\theta^TX^TX\theta - \theta^TX^TY - Y^TX\theta + Y^TY) && (2) \\
&= \frac{1}{2}\nabla_\theta tr(\theta^TX^TX\theta - \theta^TX^TY - Y^TX\theta + Y^TY) && (3) \\
&= \frac{1}{2}\nabla_\theta(tr\theta^TX^TX\theta - 2trY^TX\theta) && (4) \\
&= \frac{1}{2}(X^TX\theta + X^TX\theta - 2X^TY) && (5) \\
&= X^TX\theta - X^TY && (6)
\end{aligned}
$$

其中等式（1）类似于完全平方展开得到等式（2）；等式（2）应用$trA = A$得到等式（3）；等式（3）应用Y^TY为实数，且实数的转置为其本身，从而得到等式（4）；等式（4）应用$trABA^TC = CAB + C^TAB^T$得到等式（5）。

最后，我们令该偏导为0可得：

$$
X^TX\theta = X^TY \Rightarrow \theta = (X^TX)^{-1}X^TY
$$

从而，我们求出了参数θ的值。