Polynomial Regression

Shen Hengheng

2017

该笔记是来自 Andrew Ng 的 Machine Learning 课程的第二周: **多项式回归**的课堂记录，有了合适的特征之后，我们发现线性回归并不适用于所有数据，有时我们需要曲线来适应我们的数据，因此引出多项式回归模型，主要讲解了以下几个内容:

- 将多项式回归模型转换为线性模型
- 正规方程

0.1 将多项式回归模型转换为线性模型

什么是多项式模型？比如一个二次模型

$$h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2$$

, 三次模型

$$h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

通常情况下，在模型进行选择时，首先要观察数据，然后再决定准备尝试什么样的模型，比如下面图1 中的数据分布，我们可以通过观察可以看到，数据的分布用线性模型并不能达到很好的效果，所以在这里尝试用一个多项式模型来拟合我们的训练数据。

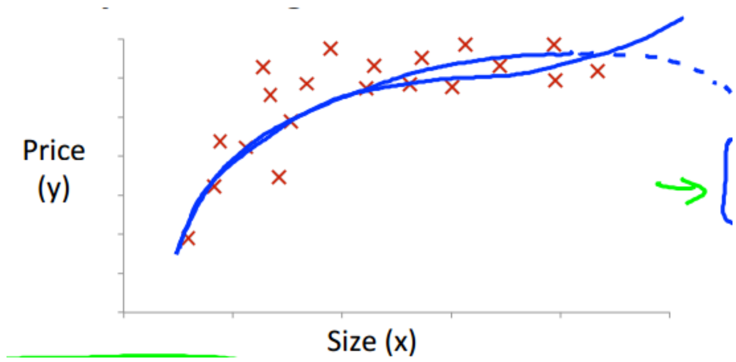

图 1: 多项式拟合

可以尝试多项式函数表示我们的模型:

$$h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2$$

或者

$$h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

或者

$$h_\theta = \theta_0 + \theta_1 x + \theta_2 x^{\frac{1}{2}}$$

有趣的是，如果我们对这些高次项进行替换，比如在 $h_\theta = \theta_0 + \theta_1 x + \theta_2 x^2$ 中，将

$$x_2 := x^2$$

将可以转换为 $h_\theta = \theta_0 + \theta_1 x + \theta_2 x_2$，这样转换为多变量线性回归模型了。
注意：在采用多项式回归模型，运行梯度下降算法前，特征缩放非常有必要！

0.2 正规方程

目前为止，都在利用梯度下降算法进行优化我们的代价函数，但对于某些的现性回归问题，正规方程是更好的解决方案。不像梯度算法那样，求解出参数 θ 需要进行很多次迭代才能求出，而正规方程直接**一次性**求出参数！其实正规方程的背后就是借助于**微积分**来做的，即通过求解下面的方程找出使得代价函数最小的参数.

$$\frac{\partial}{\partial_j} J(\theta_j) = 0$$

比如：$J(\theta) = a\theta^2 + b\theta + c$，那么通过 $\frac{d}{d_\theta} J(\theta) = ... = 0$ 可以求解出 θ 来.
经验： 使用正规方程解出向量

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

0.2.1 利用正规方程求解的例子

给你一组训练数据，$m = 4, n = 4$，则 $X \in \mathbb{R}^{4*5}, y \in \mathbb{R}^4$，数据如图2

x_0	Size (feet²) x_1	Number of bedrooms x_2	Number of floors x_3	Age of home (years) x_4	Price ($1000) y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

图 2: 数据

用矩阵可以表示为：

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix}, y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

那么就可以使用经验公式 $\theta = (x^T x)^{-1} x^T y$ 求解了！

警告：往往在使用公式求解时，存在不可逆矩阵，导致求逆矩阵失败，主要**原因**为，特征之间不独立，如同时包含英尺为单位的尺寸和米为单位的尺寸两个特征，也有可能特征数量大于训练集大于训练集数量，正规方程方法不能用！

0.2.2 梯度下降算法与正规方程的比较

梯度下降算法：

1. 不需要选择 α
2. 需要很多次迭代
3. 当特征数量 n 比较大时，也能很好的适应
4. 适应各种类型的数据

正规方程：

1. 不需要选择 α
2. 一次计算即可得出结果，不需要很多次迭代
3. 需要计算 $(x^T x)^{-1}$ 运算代价大，时间效率为 $O(n^3)$，但是如果 n<10000 时可以接受！
4. 只适应于现行模型，不适用与逻辑回归等其他类型
5. 不需要归一化特征变量