第一次讨论

名字

October 27, 2016

目　录

摘 要

上午我们讨论了线性回归，逻辑斯提回归，贝叶斯，从公式开始推导，参数的求解，以及costfunction的应用。

keywords: 线性回归; 逻辑斯提回归;贝叶斯；cost function

§1 线性回归

§1.1 一元线性回归

一元线性回归，即只有一个变量，公式的为$Y = \beta_0 + \beta_1 X$.现在我们的任务是要求解参数β_0以及β_1,这里给出两种方法：最小二乘法，梯度下降法。

§1.1.1 cost function

一元线性回归的cost function 是:

$$J(\beta_0, \beta_1) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)^2$$

最小二乘法:

分别对J关于β求导：

$$\frac{\partial J(\beta_0, \beta_1)}{\partial \beta_0} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\partial J(\beta_1, \beta_1)}{\partial \beta_0} = -\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)x_i$$

分别令上面两个式子为0，便可求得参数β_0以及β_1

梯度下降法；

分别对β_0以及β_1做偏导后，再利用梯度下降法来对参数进行更新处理，直到$\mid J(\beta_0, \beta_1)_k - J(\beta_0, \beta_1)_{k+1} \mid < \epsilon$小于某个给定的阈值。更新公式如下：

$$\beta_0 = \beta_0 - \alpha\frac{\partial J(\beta_0, \beta_1)}{\partial \beta_0} = \beta_0 - \alpha\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)$$

$$\beta_1 = \beta_1 - \alpha\frac{\partial J(\beta_0, \beta_1)}{\partial \beta_1} = \beta_1 - \alpha\frac{1}{N}\sum_{i=1}^{N}(y_i - \beta_0 - \beta_1 x_i)x_i$$

更新一次后，带入$J(\beta_0, \beta_1)$中，利用$\mid J(\beta_0, \beta_1)_k - J(\beta_0, \beta_1)_{k+1} \mid < \epsilon$

§1.2 多元线性回归

公式的形式为: $Y = \beta_0 + \beta_1 X_1 + ... + \beta_m X_m$，这里设置$X_0 = 1$,于是公式改为: $Y = \beta_0 X_0 + \beta_1 X_1 + ... + \beta_m X_m$，将公式整合到一起来写，设置$\beta = (\beta_0, \beta 1, ..., \beta(m))$,得到

$$Y = \beta^T X$$

§1.2.1 cost function

$$J(\beta) = \frac{1}{2m}\sum_{i=0}^{m}(y_i - \beta x_i)^2$$

接下来是对$J(\beta)$的每个参数β求偏导，需要注意的是$i=0$时的情形。

对于$i=0$时，

$$\frac{\partial J(\beta)}{\partial \beta_0} = \beta_0 - \alpha\frac{\partial J(\beta)}{\partial \beta_0} = \beta_0 - \alpha\frac{1}{m}\sum_{i=1}^{m}(y_i - \beta x_i)$$

对于$i>0$时，

$$\frac{\partial J(\beta)}{\partial \beta_i} = \beta_i - \alpha\frac{\partial J(\beta)}{\partial \beta_i} = \beta_i - \alpha\frac{1}{m}\sum_{i=1}^{m}(y_i - \beta x_i)x_i$$

§2 逻辑斯提回归

逻辑斯提回归是将回归的思想用于分类中。二分类中，有A和B两类。我们在使用概率时，如果$p>0.5$,会将结果判为A类；如果$p<0.5$,会将结果判为B类。

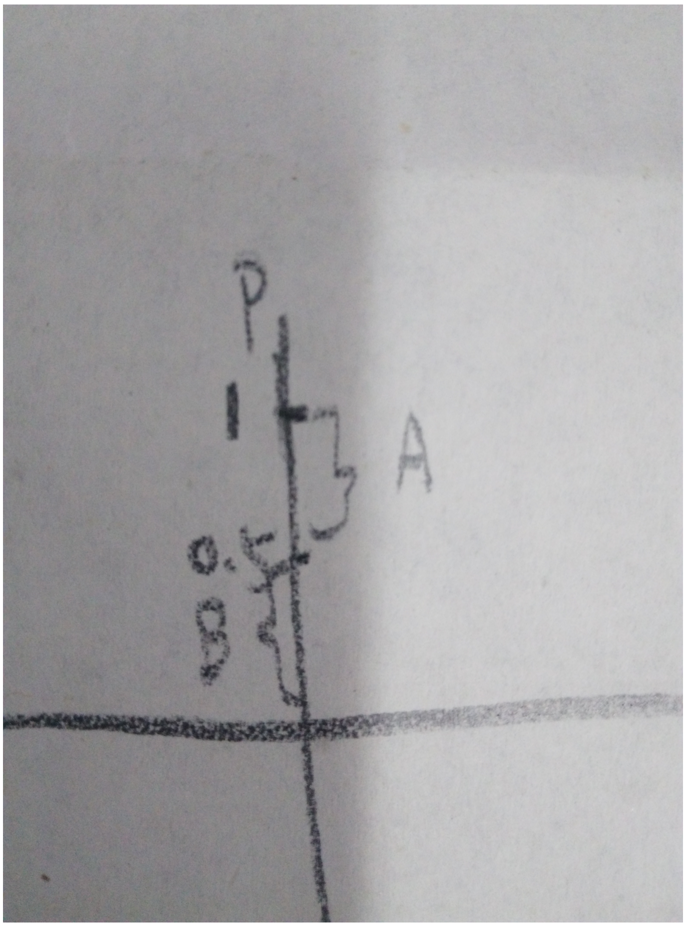

Fig. 1. 图解*1*

问题：如何引入一个平滑的函数来刻画这个过程呢？而且这个函数必须是递增的，取值范围为0到1之间。
联想到函数$\frac{1}{1+e^{-x}}$满足这个性质，图如下：

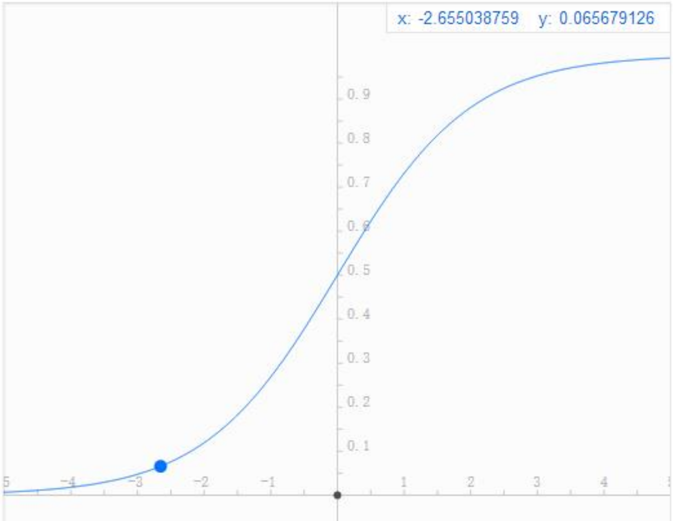

Fig. 2. 图解2

于是，得到p的函数形式，$p = \frac{1}{1+e^{-x}}$，这里的x为线性的变化，能否引入非线性的呢？当然可以，这里我们将非线性的引入，式子变为:

$$p = \frac{1}{1+e^{-a}} = \frac{1}{1+e^{-\beta^T x}}$$

对于分类问题，响应变量y是离散，没法使用cost function 来表示。回忆起我们在求解概率问题的参数时采用极大似然法来解决。同样，我们这里也采用这个方法。

$$Y = A, p = \frac{1}{1+e^{-\beta^T x}}$$

$$Y = B, p = 1 - \frac{1}{1+e^{-\beta^T x}} = \frac{e^{-\beta^T}}{1+e^{-\beta^T x}}$$

那么有：

$$J = -\frac{1}{m}[\sum_{i=0}^{m} y_i log(p) + (1-y_i)log(1-p)]$$

$$J = -\frac{1}{m}[\sum_{i=0}^{m} y_i log(\frac{1}{1+e^{-\beta^T x}}) + (1-y_i)log(\frac{e^{-\beta^T}}{1+e^{-\beta^T x}})]$$

化简后为:

$$J = -\frac{1}{m}[\sum_{i=0}^{m} y_i \beta^T x - log(1+e^{\beta^T x})]$$

§3 贝叶斯

贝叶斯的使用假设是：数据是独立同分布的。

将贝叶斯用于分类中，以概率为判断准则，谁的概率值大，结果就属于对应的

类别。

公式:

$$p(Y|X) = \frac{p(Y)p(X|Y)}{p(X)}$$

这里的$p(X)$、$p(Y)$属于先验信息。$p(X|Y)$属于条件概率。通过这三个量来求后验概率，从而达到分类的效果。又由于对某个数据，需要将它分类，在所有类的判别时，它的概率是不变的，即分母不变。公式还可以改为:

$$p = p(Y)p(X|Y) = p(Y)p(x^1, x^2, ..., x^m|Y = c_k) = p(Y)\prod_{i=1}^{m} p(x^i|Y = c_i)$$

§3.1 离散型的

上面式子的$p(X)$、$p(Y)$、$p(X|Y)$是可以直接计算出来，从而求出后验信息$p(Y|X)$.

§3.2 连续型的

根据离散型的可知，得到连续型的形式为:

$$p = p(Y)f$$

这里的f为数据x所服从分布的密度函数。这个在LAD和QDA中将使用到。