线性回归模型

一、线性回归模型定义

1.它是用来做什么的?

本质上是**通过给定的特征X，预测目标y值**的一个模型，其中的X是一个向量，表示可以有多个特征。

2.它的数学表达式是什么?

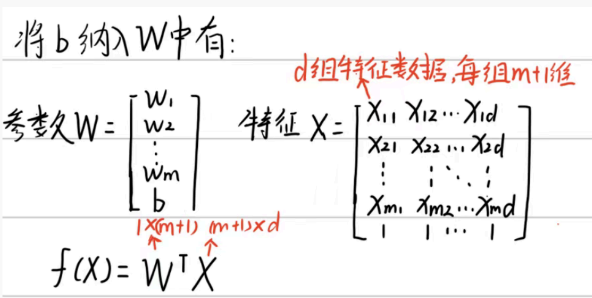
$$f(x) = w^T x + b \tag{1}$$

其中，w和x均为向量，通过公式可以看出线性回归模型就是要**求得一组最优的wi和b**来确定线性模型，使这个模型**无限逼近**现有的数据xi和结果$f\ xi$ 之间的关系。

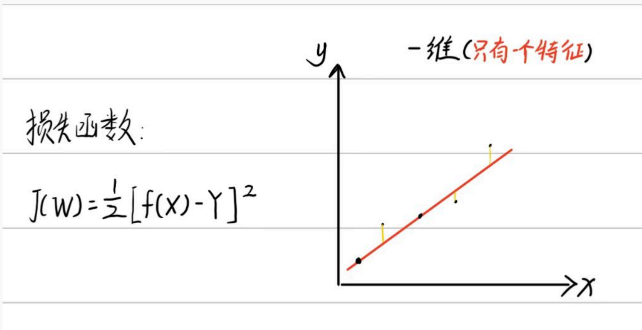
二、损失函数

我们的目标是求解最优参数使得模型无限逼近真实关系，因此为了评判逼近程度，我们定义一个**损失函数（用它来描述模型计算输出与真实值之间的误差）**，当损失函数最小时，就表明最逼近现有关系。

为了计算方便，将（1）式的b纳入矩阵W中，得到线性回归模型的向量表达式如（a）图所示：



(a)线性回归模型向量表示



(b)损失函数向量表示

根据损失函数的定义可以得到其表达式如（b）图左侧所示，

$$J(W) = \frac{1}{2}[W^T X - Y]^2 \tag{2}$$

（b）图右侧是一维时的示意图，其中红色的线表示求出的线性回归模型，黄色的线表示误差，**损失函数就是所有误差平方的和**，平方是为了保证损失度恒为正，1/2是方便求导。

三、参数求解方式

1.最小二乘法

我们需要找到最优的参数使得损失函数最小，从数学的角度看就是**对损失函数求导，当导函数为0时取得最小值**，最小二乘法就是用这个方式求解参数的值，下面给出具体的推导：

损失函数: $J(W)=\frac{1}{2}[f(X)-Y]^2 \quad f(X)=W^TX$

$=\frac{1}{2}(W^TX-Y)^2$

$=\frac{1}{2}(W^TX-Y)^T(W^TX-Y)$ $\quad (W^TX)^T=X^TW$

$=\frac{1}{2}(X^TW-Y^T)(W^TX-Y)$

$=\frac{1}{2}(X^TWW^TX-Y^TW^TX-X^TWY+Y^TY)$

$Y^TW^TX=X^TWY$

$=\frac{1}{2}(W^TXX^TW-2X^TWY+Y^TY)$

矩阵求导的三个公式: ① $\frac{\partial AB}{\partial B}=A^T$ ② $\frac{\partial A^TB}{\partial A}=B$ ③ $\frac{\partial X^TAX}{\partial X}=2AX$

对损失函数求导:

$\frac{\partial J(W)}{\partial W}=\frac{1}{2}\left[\frac{\partial(W^TXX^TW)}{\partial W}-2\frac{\partial(2Y^TXW)}{\partial W}\right]$

$=\frac{1}{2}(2XX^TW-2X^TY)$

$=XX^TW-X^TY$

令导为0 $\quad XX^TW-X^TY=0$ 则 $W=(XX^T)^{-1}X^TY$

最终得到最优参数:

$$W=(XX^T)^{-1}X^TY \tag{3}$$

可以看到由于逆函数不一定存在，因此存在无法求出W最优W的局限性。

2.梯度下降法

梯度下降法通俗讲就是**沿着损失函数梯度的反方向更新参数使得损失函数不断趋近最小值**，具体公式如下所示:

$$W=W-\alpha\frac{\partial J(W)}{\partial W} \tag{4}$$

其中α表示每一步的大小，过大可能导致在最小值两侧振荡，过小会导致趋近速度太慢。

具体的伪代码步骤:

1. 初始化$W=W_0$
2. 循环更新迭代$W=W-\alpha\frac{\partial J(W)}{\partial W}$
3. 循环终止条件：损失函数与0的误差满足需求（例如$J(W)-0<10^{-5}$）