

样本数据[（**X1**，**Y1**）, （**X2**，**Y2**）, …(**XM,YM**) ]，有M张图片样本，（X1，Y1）代表第一张图片，X1= {x 11，x 12，x 13….. x 1d}表示图片1的1-d个像素值，Y1{y 11，y 12，y 13}是图片1的标签，表示这张图片是啥：猫（{1,0,0}）, 狗（{0,1,0}）, 人（{0,0,1}）

X1·W1 = B1即{b1…bq}

B1’ = 激活函数（B1）

* **sigmoid函数：**
* **Tanh函数：**
* **ReLU**

B1’·W2 = Y1’，这里的Y1’表示预测的标签

那么如何判断Y1’预测的对不对呢？ 看Y1’和Y1 的差别大不大

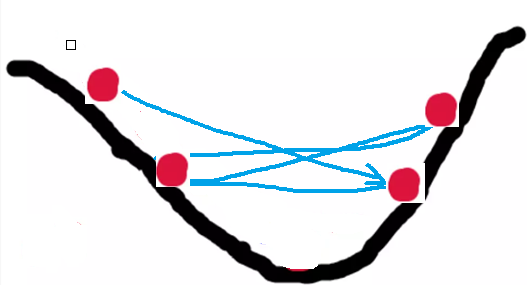
引入“损失函数”的概念

= f(W),是一个w的函数，那我目标就是求L取得最小值时的W值。

由梯度下降法可知，w = w –η\* dw η为学习率



学习率没选择好，模型就无法收敛。



损失函数改进 = 后面这块就叫做正则化项

Cost = Loss的累加，叫做代价函数，把所有M个样本的损失函数累加起来，一起训练。

## 学习的回合数、迭代次数：epoch，iteration；

1个epoch等于使用训练集中的全部样本训练一次；：1个iteration等于使用batchsize个样本训练一次。

## 小批量数据 minibatch的大小；

批大小。在[深度学习](https://www.baidu.com/s?wd=%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1Y3nvf4nHubrjI-njDkuHNB0ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EnW0krHRkn1c3P1R4PWbkPWfd)中，一般采用SGD训练，即每次训练在训练集中取batchsize个样本训练。

一次计算所有的cost效率太低，收敛难度大，所有，需要分多批次来训练，每批次，降低一点Cost。

