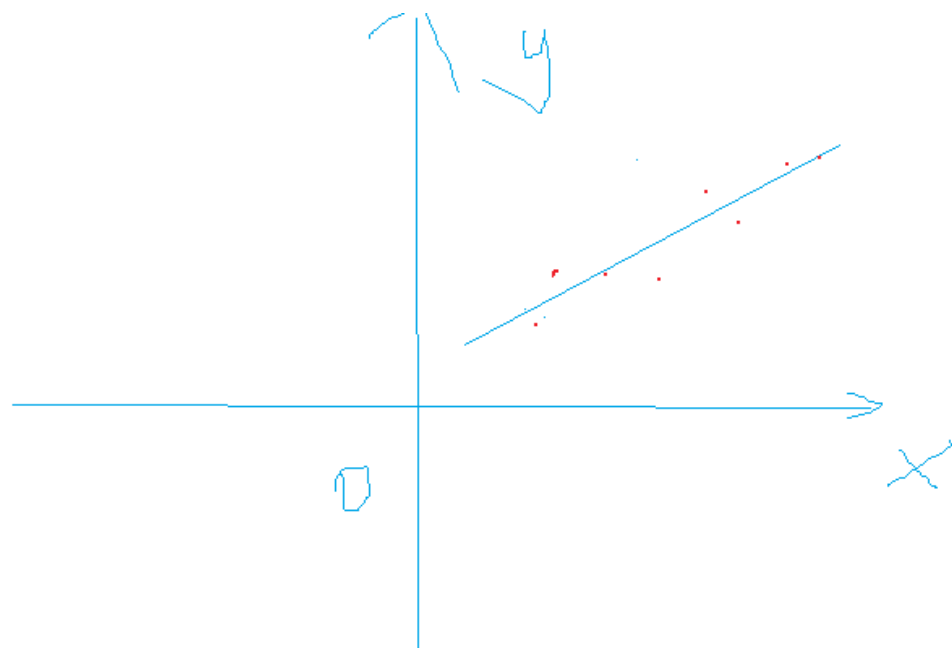


深入了解机器学习（线性回归和损失）

● 线性回归

线性关系想必大家都很熟悉， $y = k \cdot x + b$ （ k 为常数， b 为常数）

Y 可以代表我们要预测的值， $m = 1/k$ （即直线的斜率）， x 代表已经知道的值（即特征值，可以当做某一因素）， b 代表 y 轴的截距。简易图如下图所示：



红色的点代表已经发生的事件，这条直线的可以近似地来表示

这种关系（某事件受某种因素而发生的关系）。我们可以按照上述关系写一个细微差别的模型方程式：

$$Y' = b + w1 * x1 / Y' = b + w1 * x1 + w2 * x2 (\text{多个特征})$$

Y' : 预测标签即理想输出值（标签）

b : 偏差

$w1$: 指的是特征值的权重，权重和线性方程中的斜率概率相同。

$x1$: 指的是特征（已知输入项）

● 训练与损失

■ 训练模型表示通过标签样本来学校确定所有权重和偏差的理想值。

■ 损失是指对糟糕预测的惩罚，会影响模型预测的准确程度，预测完全准确，损失为零。

平方损失：一种常见的损失函数，单个样本的平方损失可用下面表达式来表示：

$$(\text{observation} - \text{prediction}(x))^2 = (y - y')^2$$

均方损失：是指每个样本的平方损失，表达式关系如下：

$$MSE = \frac{1}{N} \sum_{(x,y) \in D} (y - \text{prediction}(x))^2$$

(x,y) : 样本， x （特征值）， y （样本的标签）。 $\text{Prediction}(x)$ （权重和偏差与特征集 x 结合的函数）， D 指包含多个样本的集合，

N 代表 D 中的样本数量。