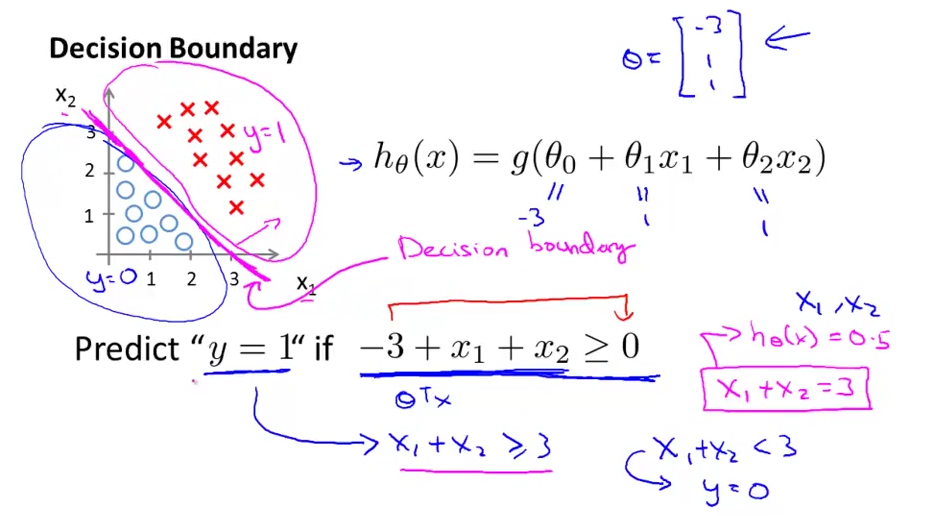
**Decision Boundary**

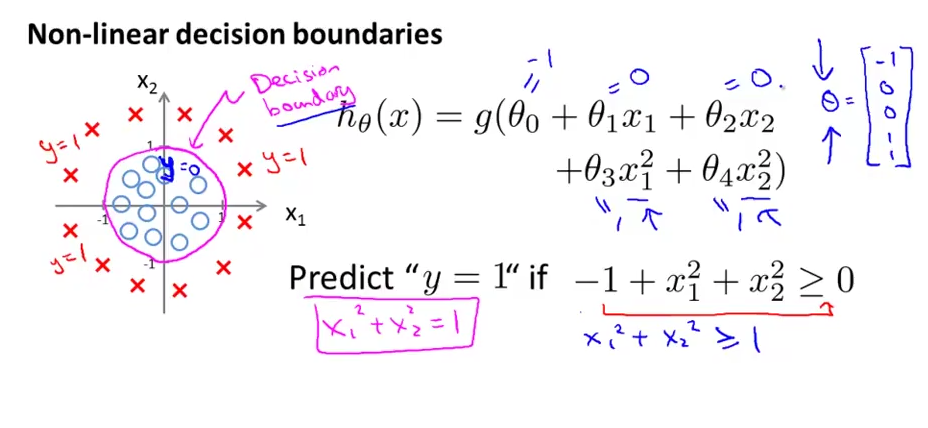
决策边界就是能够把样本正确分类的一条边界，主要有线性决策边界(linear decision boundaries)和非线性决策边界(non-linear decision boundaries)。注意：决策边界是假设函数的属性，由参数http://img.blog.csdn.net/20180124165741064决定，而不是由数据集的特征决定。下面主要举一些例子，形象化的来说明线性决策边界和非线性决策边界。

先看一个线性决策边界的例子：（注：图片来源：ng的machine learning课）

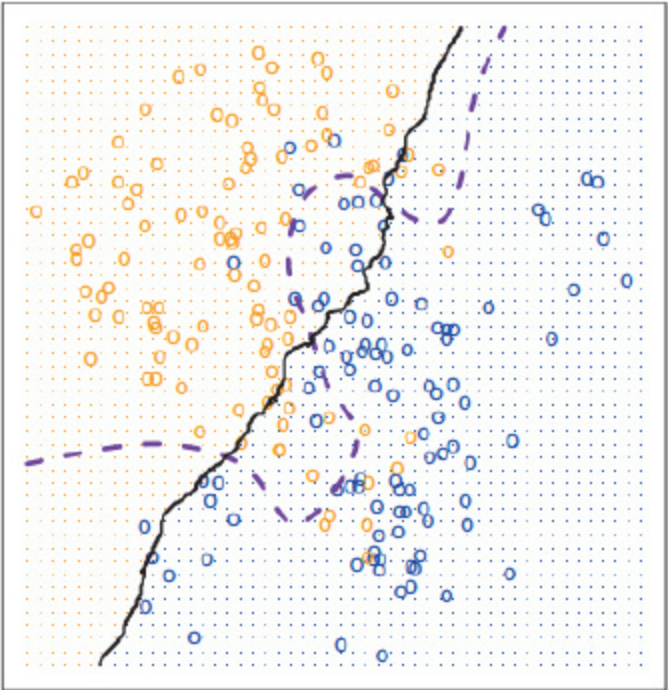


再来看一个非线性决策边界的例子：

假设我们的数据呈现出如下图的分布情况，那么我们的模型是什么样才能适合这些数据呢？我们需要的是一个二次方特征。



由不同机器学习算法可以得到不同的决策边界，不同的数据或不同的算法都会影响边界的准确性。以下图为例，紫色虚线是贝叶斯决策边界线，黑色实线则是KNN的分类边界，由图可以发现对于该组数据贝叶斯决策边界线是较理想的分类边界。



参考资料：<http://blog.csdn.net/u012328159/article/details/51068427>

**F1 Score**

F1分数（F1 Score），是统计学中用来衡量二分类模型精确度的一种指标。它同时兼顾了分类模型的[准确率](https://baike.baidu.com/item/%E5%87%86%E7%A1%AE%E7%8E%87)和[召回率](https://baike.baidu.com/item/%E5%8F%AC%E5%9B%9E%E7%8E%87)。F1分数可以看作是模型[准确率](https://baike.baidu.com/item/%E5%87%86%E7%A1%AE%E7%8E%87)和[召回率](https://baike.baidu.com/item/%E5%8F%AC%E5%9B%9E%E7%8E%87)的一种加权平均，它的最大值是1，最小值是0。

* 数学定义

F1分数（F1Score），又称平衡F分数（balanced F Score），它被定义为精确率和[召回率](https://baike.baidu.com/item/%E5%8F%AC%E5%9B%9E%E7%8E%87)的[调和平均数](https://baike.baidu.com/item/%E8%B0%83%E5%92%8C%E5%B9%B3%E5%9D%87%E6%95%B0)。

https://gss3.bdstatic.com/7Po3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D182/sign=eed1ed8a5e6034a82de2bc89f91149d9/4610b912c8fcc3ceb409adb89045d688d53f206b.jpg

更一般的，我们定义FB分数为

https://gss1.bdstatic.com/9vo3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D265/sign=30335b98df54564ee165e33f86df9cde/c2cec3fdfc039245afef86848594a4c27c1e25e6.jpg

* 物理意义

人们通常使用准确率和召回率这两个指标，来评价二分类模型的分析效果。但是当这两个指标发生冲突时，我们很难在模型之间进行比较。比如，我们有如下两个模型A、B，A模型的召回率高于B模型，但是B模型的准确率高于A模型，A和B这两个模型的综合性能，哪一个更优呢？

|  |  |  |
| --- | --- | --- |
| 模型 | 准确率 | 召回率 |
| A | 80% | 90% |
| B | 90% | 80% |

为了解决这个问题，人们提出了FB分数。FB的物理意义就是将准确率和召回率这两个分值合并为一个分值，在合并的过程中，召回率的权重是准确率的B倍。F1分数认为召回率和准确率同等重要，F2分数认为召回率的重要程度是准确率的2倍，而F0.5分数认为召回率的重要程度是准确率的一半。

* 应用领域

F分数被广泛应用在[信息检索](https://baike.baidu.com/item/%E4%BF%A1%E6%81%AF%E6%A3%80%E7%B4%A2/831904)领域，用来衡量检索分类和文档分类的性能。早期人们只关注F1分数，但是随着谷歌、百度等大型搜索引擎的兴起，召回率和准确率对性能影响的权重开始变得不同，人们开始更关注其中的一种，所以FB分数得到越来越广泛的应用。F分数也被广泛应用在自然语言处理领域，比如命名实体识别、分词等，用来衡量算法或系统的性能。

補充：

1. G分数是另一种统一[准确率](https://baike.baidu.com/item/%E5%87%86%E7%A1%AE%E7%8E%87)和[召回率](https://baike.baidu.com/item/%E5%8F%AC%E5%9B%9E%E7%8E%87)的系统性能评估标准。F分数是准确率和召回率的[调和平均数](https://baike.baidu.com/item/%E8%B0%83%E5%92%8C%E5%B9%B3%E5%9D%87%E6%95%B0)，G分数被定义为准确率和召回率的[几何平均数](https://baike.baidu.com/item/%E5%87%A0%E4%BD%95%E5%B9%B3%E5%9D%87%E6%95%B0)。

https://gss3.bdstatic.com/7Po3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D162/sign=c20d5b523b01213fcb334ada66e636f8/d058ccbf6c81800a56c5e685b33533fa828b471c.jpg

2. 精确率、召回率、F1 值、ROC、AUC 各自的优缺点是什么?

<https://www.zhihu.com/question/30643044>

参考资料：<https://baike.baidu.com/item/F1%E5%88%86%E6%95%B0/13864979?fr=aladdin>

**Partial Derivative**

在数学中，一个多变量的函数的偏导数，就是保持其他变量恒定下关于其中一个变量的导数而（相对于全导数，在其中所有变量都允许变化）。偏导数在向量分析和微分几何中用途广泛。

* 引入

在一元函数中，导数就是函数的变化率。对于二元函数研究它的“变化率”，由于自变量多了一个，情况就要复杂的多。

在 xOy 平面内，当动点由 P(x0,y0) 沿不同方向变化时，函数 f(x,y) 的变化快慢一般说来是不同的，因此就需要研究 f(x,y) 在 (x0,y0) 点处沿不同方向的变化率。

在这里我们只学习函数 f(x,y) 沿着平行于 x 轴和平行于 y 轴两个特殊方位变动时， f(x,y) 的变化率。

偏导数的表示符号为:∂。

偏导数反映的是函数沿坐标轴正方向的变化率。

* 定义

x方向的偏导

设有二元函数 z=f(x,y) ，点(x0,y0)是其定义域D 内一点。把 y 固定在 y0而让 x 在 x0 有增量 △x ，相应地函数 z=f(x,y) 有增量（称为对 x 的偏增量）△z=f(x0+△x,y0)-f(x0,y0)。

如果 △z 与 △x 之比当 △x→0 时的极限存在，那么此极限值称为函数 z=f(x,y) 在 (x0,y0)处对 x 的偏导数，记作 f'x(x0,y0)。函数 z=f(x,y) 在(x0,y0)处对 x 的偏导数，实际上就是把 y 固定在 y0看成常数后，一元函数z=f(x,y0)在 x0处的导数。

y方向的偏导

同样，把 x 固定在 x0，让 y 有增量 △y ，如果极限存在那么此极限称为函数 z=(x,y) 在 (x0,y0)处对 y 的偏导数。记作f'y(x0,y0)。

* 求法

当函数 z=f(x,y) 在 (x0,y0)的两个偏导数 f'x(x0,y0) 与 f'y(x0,y0)都存在时，我们称 f(x,y) 在 (x0,y0)处可导。如果函数 f(x,y) 在域 D 的每一点均可导，那么称函数 f(x,y) 在域 D 可导。

此时，对应于域 D 的每一点 (x,y) ，必有一个对 x (对 y )的偏导数，因而在域 D 确定了一个新的二元函数，称为 f(x,y) 对 x (对 y )的偏导函数。简称偏导数。

按偏导数的定义，将多元函数关于一个自变量求偏导数时，就将其余的自变量看成常数，此时他的求导方法与一元函数导数的求法是一样的。

* 几何意义

表示固定面上一点的切线斜率。

偏导数 f'x(x0,y0) 表示固定面上一点对 x 轴的切线斜率；偏导数 f'y(x0,y0) 表示固定面上一点对 y 轴的切线斜率。

高阶偏导数：如果二元函数 z=f(x,y) 的偏导数 f'x(x,y) 与 f'y(x,y) 仍然可导，那么这两个偏导函数的偏导数称为 z=f(x,y) 的二阶偏导数。二元函数的二阶偏导数有四个：f"xx，f"xy，f"yx，f"yy。

注意：

f"xy与f"yx的区别在于：前者是先对 x 求偏导，然后将所得的偏导函数再对 y 求偏导；后者是先对 y 求偏导再对 x 求偏导。当 f"xy 与 f"yx 都连续时，求导的结果与先后次序无关。

补充：

如何理解导数的概念 ? <https://www.zhihu.com/question/28684811>

参考资料：<https://baike.baidu.com/item/%E5%81%8F%E5%AF%BC%E6%95%B0/5536984?fr=aladdin>