机器学习算法

# 感知器

模型：

输入空间,输出空间, 对每一个输入表示一个实例，输出表示实例的类别，由输入空间到输出空间有如下的函数：

学习策略：

1、分类错误点数最少；

2、误分类点到超平面的距离之和最小。

学习算法:

梯度下降算法

# K近邻

模型：

输入训练集：T = {(x1,y1) , …..(xn, yn)}其中xi∈为实例的特征向量，yi∈为实例的类别，i = 1..N; 实例特征向量x

输出：实例x的类别为y

1. 根据给定的距离度量，在训练集中找出与x最邻近的k个点，涵盖着k个点的x的领域记为Nk(x)
2. 在Nk(x)中根据分类决策规则（如多数表决）决定x的类别y:

模型的关键是：确定距离度量方式，k值的大小以及分类决策的规则。

距离度量：使用向量的p范数，作为距离度量的方式

k值：需要根据具体的情况来选择，其影响着模型的复杂度

分类决策规则：多数表决规则。

误分类概率：P(Y≠f(X)) = 1- P(Y = f(X))

那么在 )邻域内，若其类别为cj, 那么误分类率为：

, 即误分类率最小，也就是经验风险最小。

构造kd树：