第一章：绪论

1、正确分类与训练集不同的新样本的能⼒叫做泛化，所以泛化是模式识别的⼀个中⼼问题。2、训练数据的样本包含输⼊向量以及对应的⽬标向量的应⽤叫做有监督学习问题。数字识别就是这个问题的⼀个例⼦，它的⽬标是给每个输⼊向量分配到有限数量离散标签中的⼀个，被称为分类问题。如果要求的输出由⼀个或者多个连续变量组成，那么这个任务被称为回归。回归问题的⼀个例⼦是化学药品制造过程中产量的预测。在这个问题中，输⼊由反应物、温度、压⼒组成。

3、训练数据由⼀组输⼊向量x组成，没有任何对应的⽬标值。⽆监督学习，⽬标可能是发现数据中相似样本的分组，这被称为聚类或者决定输⼊空间中数据的分布，这被称为密度估计，或者把数据从⾼维空间投影到⼆维或者三维空间，为了数据可视化。4、多项式曲线拟合是一个回归问题：

（1）虽然多项式函数y(x, w)是x的⼀个⾮线性函数，它是系数w的⼀个线性函数

（2）由于误差函数是系数w的⼆次函数，因此它关于系数的导数是w的线性函数，所以误差函数的最⼩值有⼀个唯⼀解，记作w∗，可以⽤解析的⽅式求出。

（3）事实上，多项式函数精确地通过了每⼀个数据点，E(w∗) = 0。然⽽，拟合的曲线剧烈震荡，就表达函数sin(2πx)⽽⾔表现很差。这种⾏为叫做过拟合（over-fitting）