# 大数据与机器学习之旅    
### 基本概念    
#### 【模式识别与机器学习】模式识别中的一些基本概念    
\* 特征(feature)    
```  
如果有一个区分鱼的类别的系统，可以分类的依据为长度、光泽、宽度、鳍的数目和形状、嘴的位置。这些可以利用的要素称为模式分类的特征。  
```  
  
\* 模型(model)    
```  
如果鱼的不同类别之间确实存在某种差异，我们称之为具有不同的模型，即可以用数学形式表达的不同特征的描述。    
在模式识别系统中，经常会用到三种密切相关的系统，回归分析、函数内插和（概率）密度估计。    
```  
  
\* 回归(regression)    
```  
回归分析中，我们的目的是对输入数据找到合适的函数表示，常用于预测新数据的值。线性回归(linear regression)，其中的函数形式对输入数据而言是线性的，是目前研究的最透彻的一种回归形式。    
```  
  
\* 函数内插(interpolation)    
```  
函数内插中，我们已知的是一定范围内的输入数据对应的函数值，而要解决的问题是如何求出位于这些输入点之间的数据点的函数值。间隔的数据需要内插技术来解决。    
```  
  
\* 密度函数估计(density estimation)  
```  
密度函数估计用于求解具有某种特定特征的类别成员(样本)出现的(概率)密度的问题。    
```  
  
\* 先验概率    
```  
假设捕捉到的鲈鱼和鲑鱼的数目是相等的，那么可以说下一次出现鲈鱼和出现鲑鱼的可能性是相等的。一般的，假设下一条鱼的先验概率为p(w1)，而下一条鱼是鲑鱼的先验概率为p(w2)。由于没有其他类别的鱼，所以又p(w1) + p(w2) = 1。如果，在做出鱼的种类判别时，没有其他的判别信息，那么可以利用的只有先验概率。    
```  
  
\* 类条件概率密度(class-conditional probability density)函数    
如果可以利用的判别信息有光泽度指标为x，不同的鱼有不同的光泽度。将其表示成概率形式的变量，假定x是一个连续随机变量，其分布取决于类别的状态，表示成p(x|w)的形式，这就是类条件概率密度，即类别状态为w时的x的概率密度函数。于是p(x|w1)与p(x|w2)之间的区别就成了表示鲈鱼与鲑鱼间光泽度的区别。  
  
#### 机器学习：强化学习