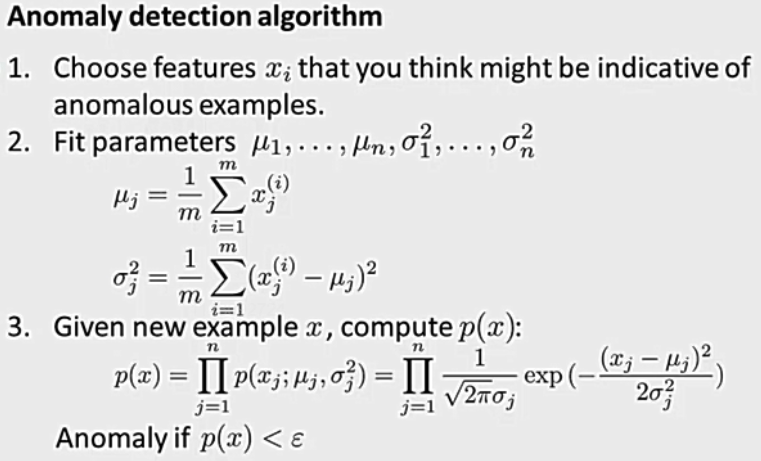
1、异常检测

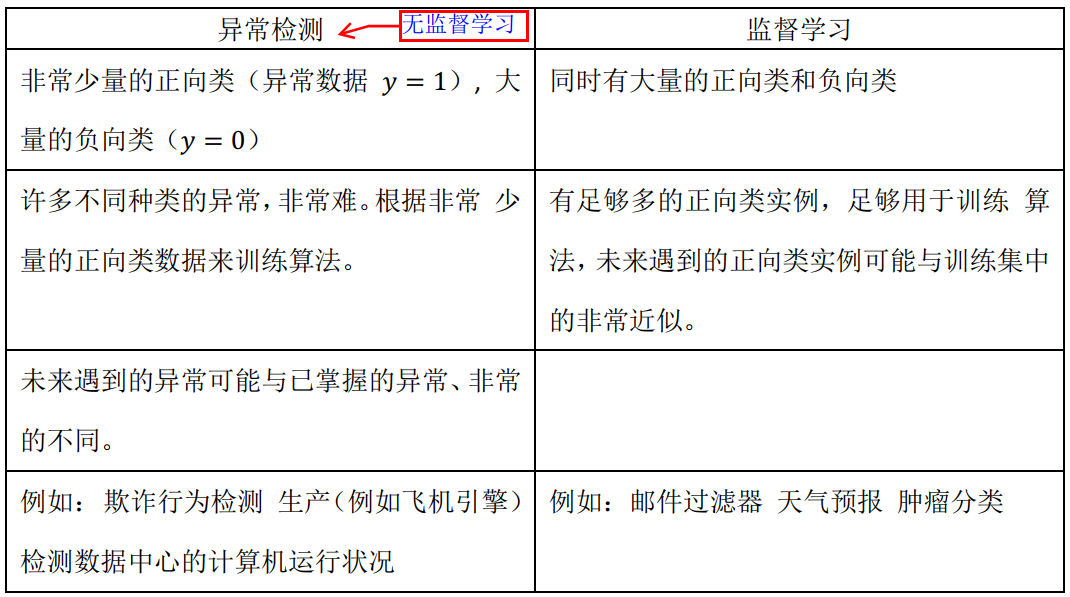
（1）概述



使用高斯公式来判断，当概率小于时，可以认定该数据为异常点。

通常都比较小，eg：

（2）区别

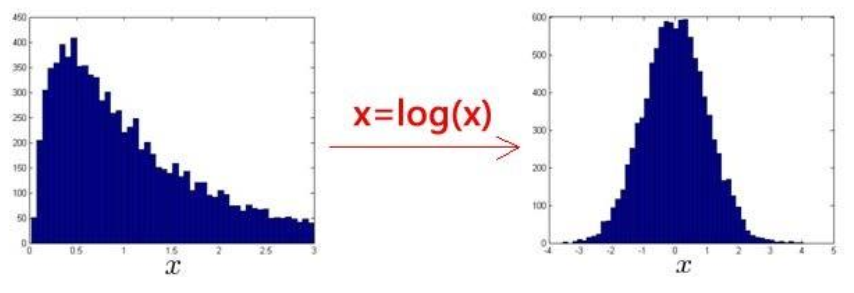


（3）数据处理

若数据没有符合高斯分布，最好还是转换成高斯分布。

例如，使用对数函数 ，其中c为非负常数；

或者，c为 0-1 之间的一个分数。



2、多元高斯分布：（二维变量）正相关、负相关

1）描述两个特征变量之间的关系。

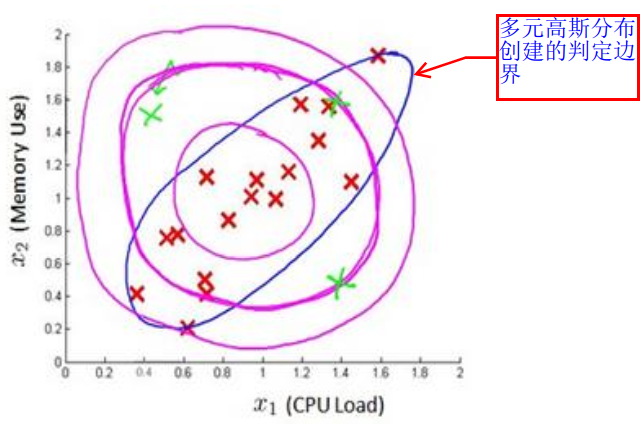
2）通过改变均值、协方差，来改变分布状态。

3）使用多元高斯分布：

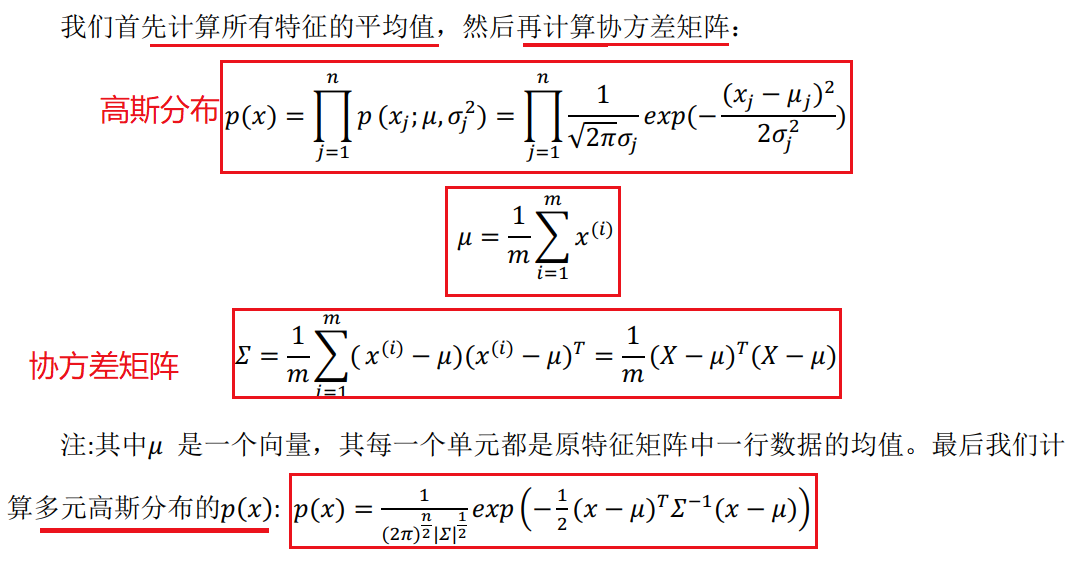
1、需要使数据数量m>10倍的特征数量（n）

即：m>10n

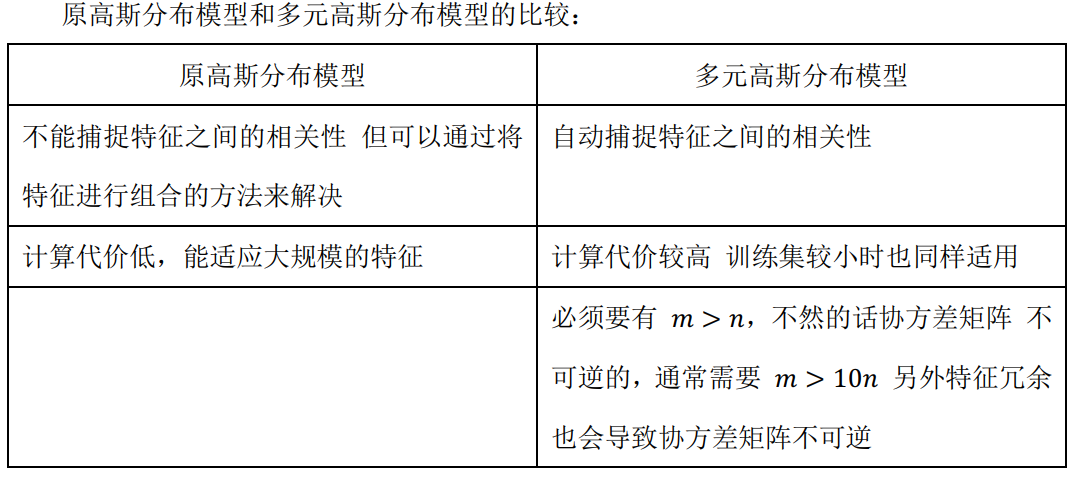
2、去除数据冗余



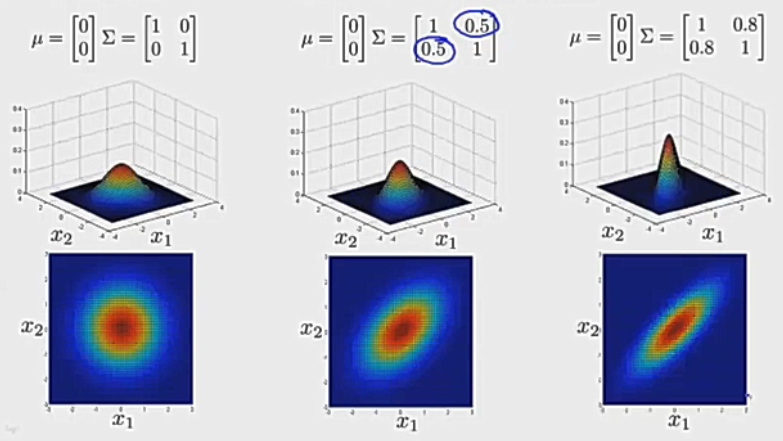
3、多元高斯分布计算步骤：

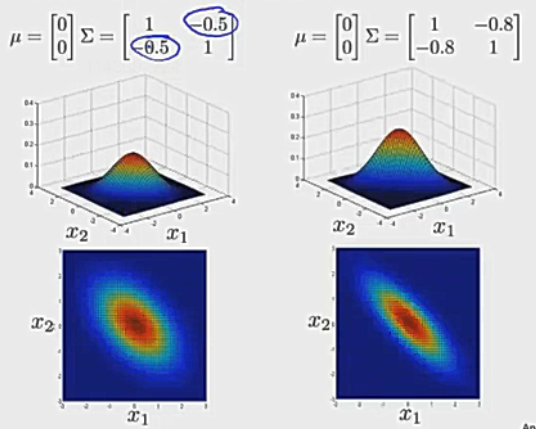


4、区别



4）高斯分布图像

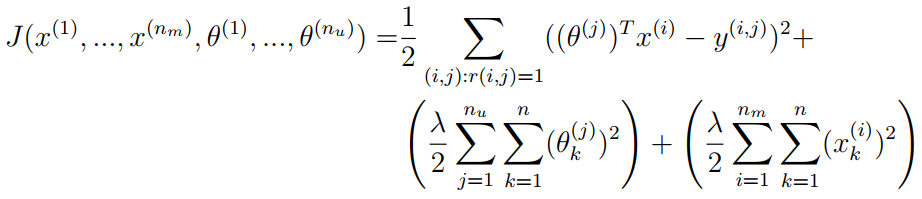




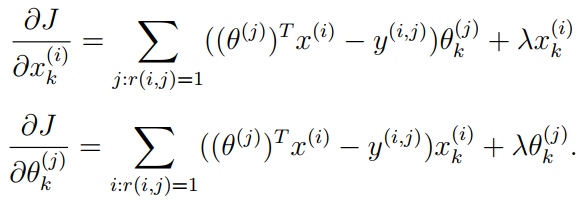
3、推荐系统

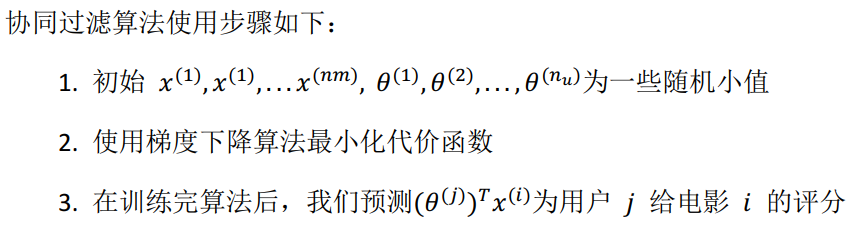
（1）协同过滤

1、代价函数

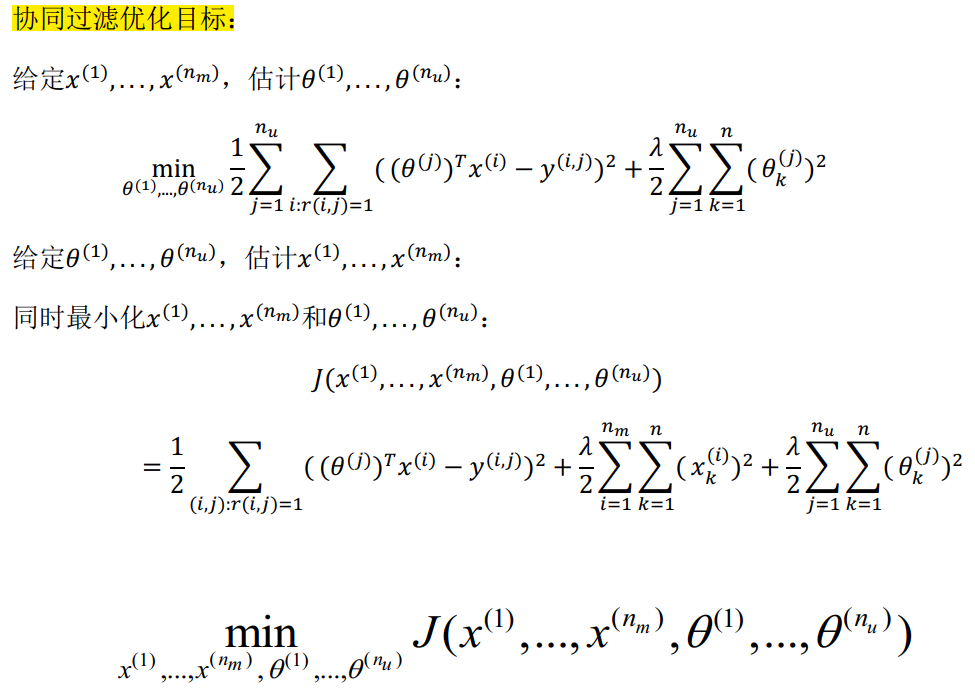


2、偏导数





（2）协同过滤算法



（2）矢量化：低秩矩阵分解

（3）均值规范化

