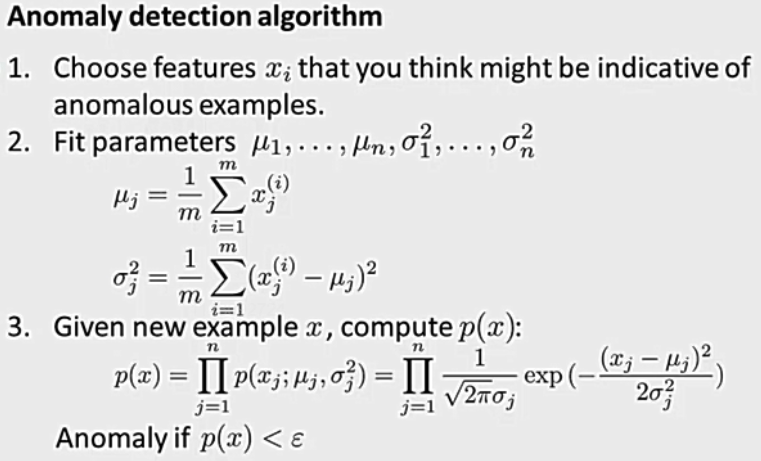
1、异常检测

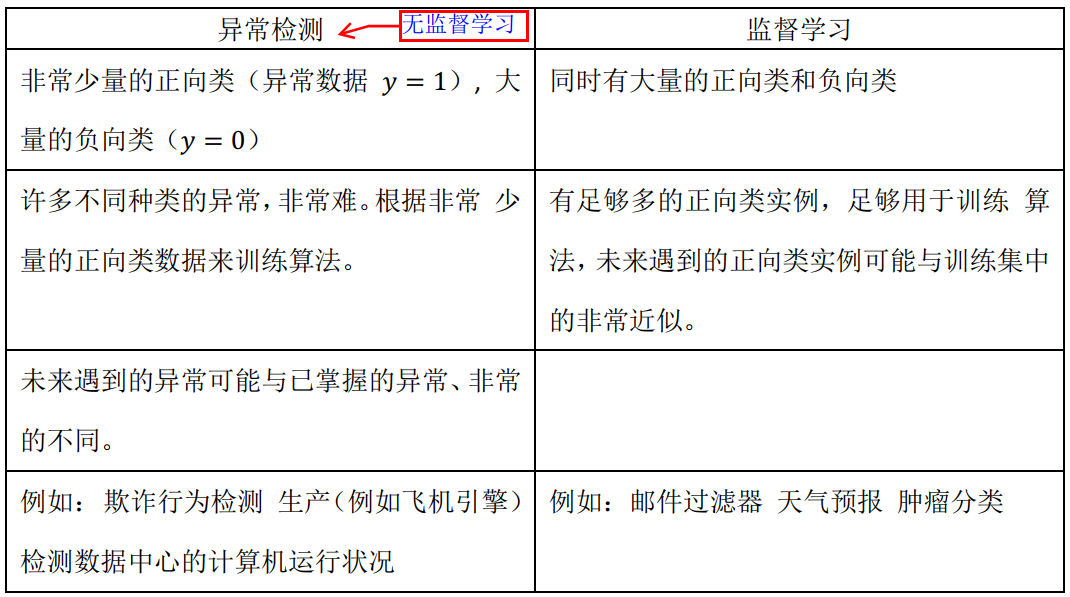
（1）概述



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

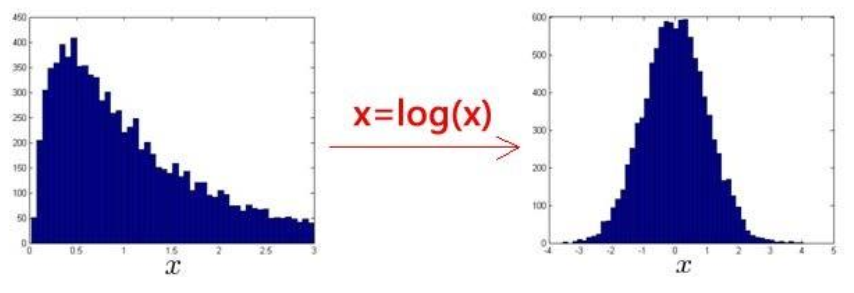


（3）数据处理

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。

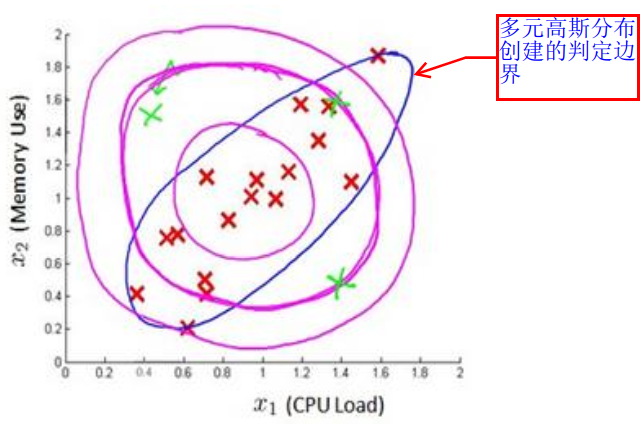
2）通过改变均值、协方差，来改变分布状态。

3）使用多元高斯分布：

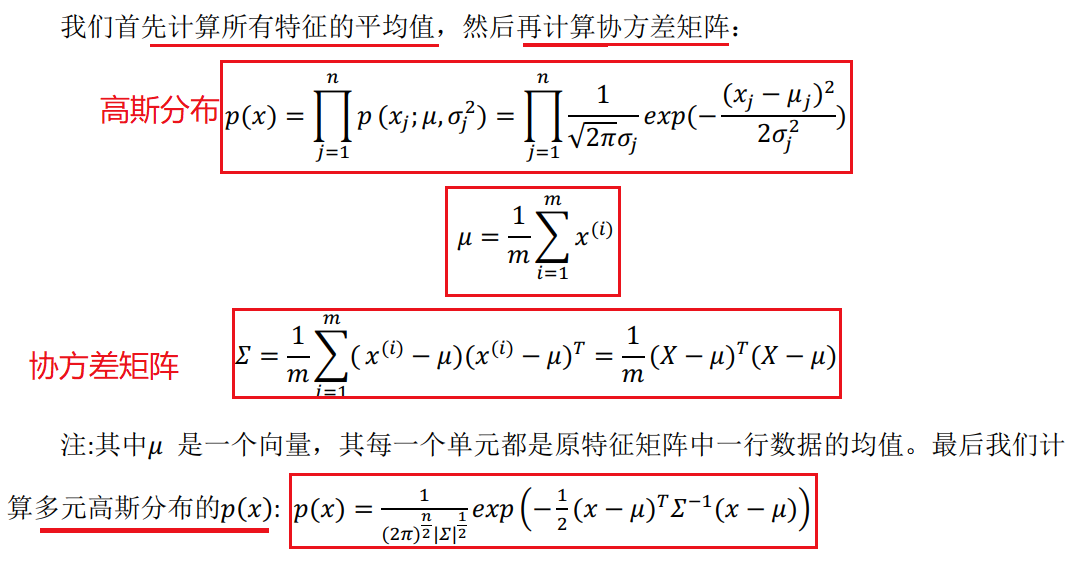
1、需要使数据数量m>10倍的特征数量（n）

即：m>10n

2、去除数据冗余

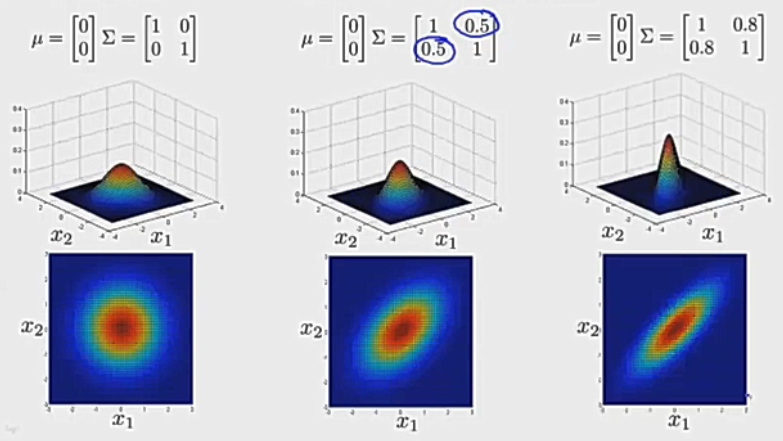


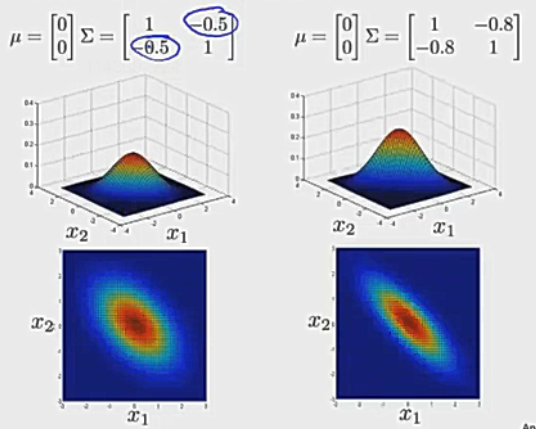
3、多元高斯分布计算步骤：



4、区别

4）高斯分布图像





3、推荐系统

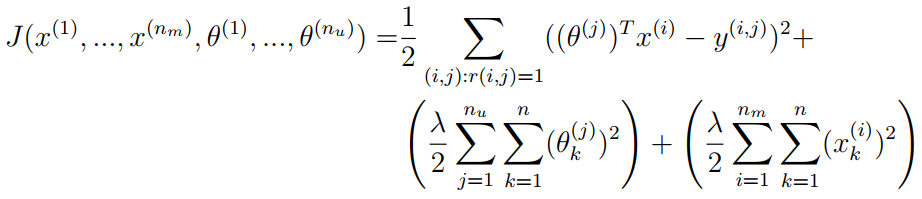
（1）协同过滤

**思想：**通过将用户和其他用户的数据进行对比来实现推荐。

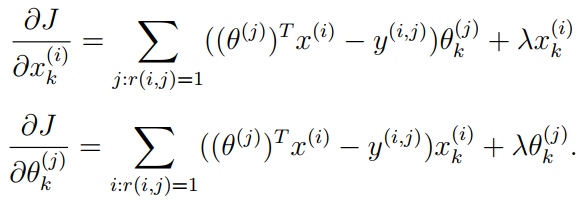
**核心：**相似度计算方法（见：笔记06，补充知识）

**协同：**的意思——通过用户对电影的评分，来帮助系统学习参数theta，然后使用习得的参数theta来评估用户的分数……以此反复循环，使得theta达到最佳。

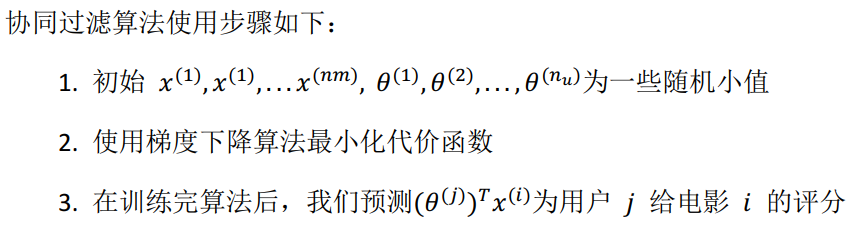
1、代价函数

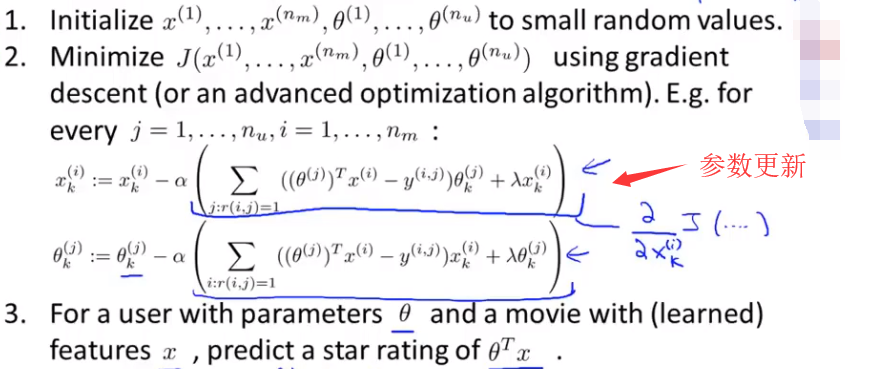


2、偏导数

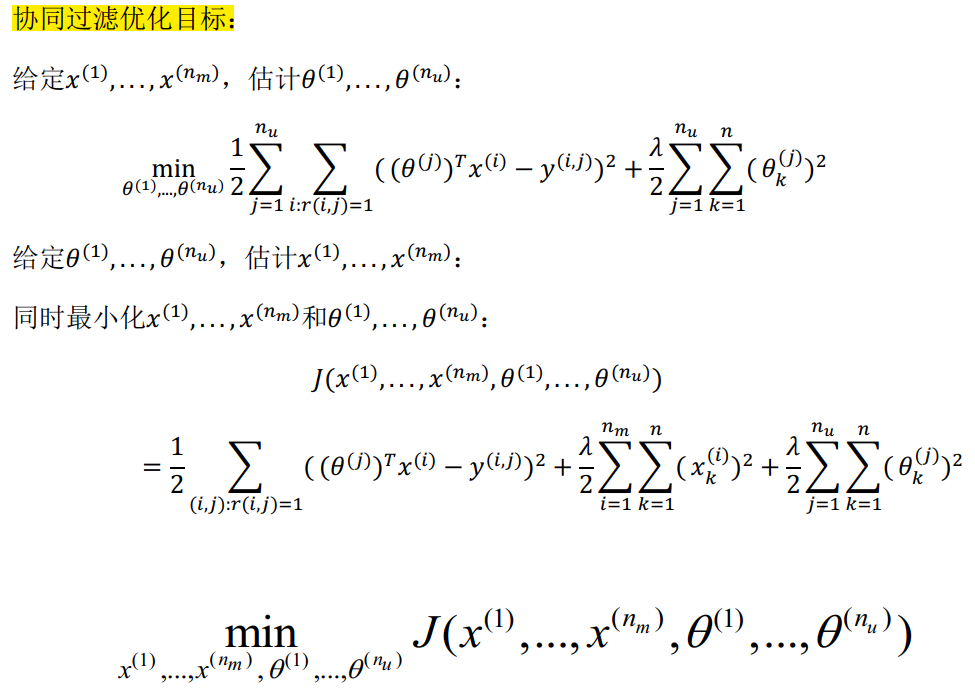


此处的**r(i,j)=1**，表示**用户j**评价过的**电影i**

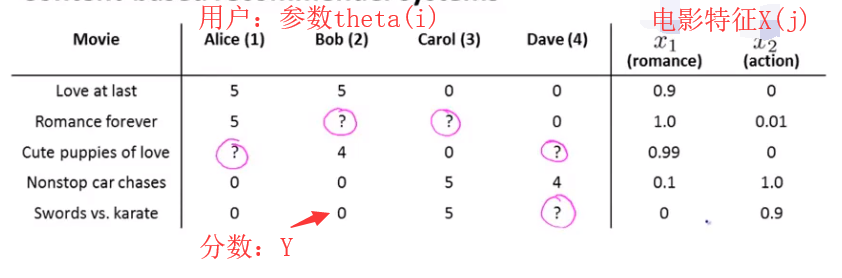


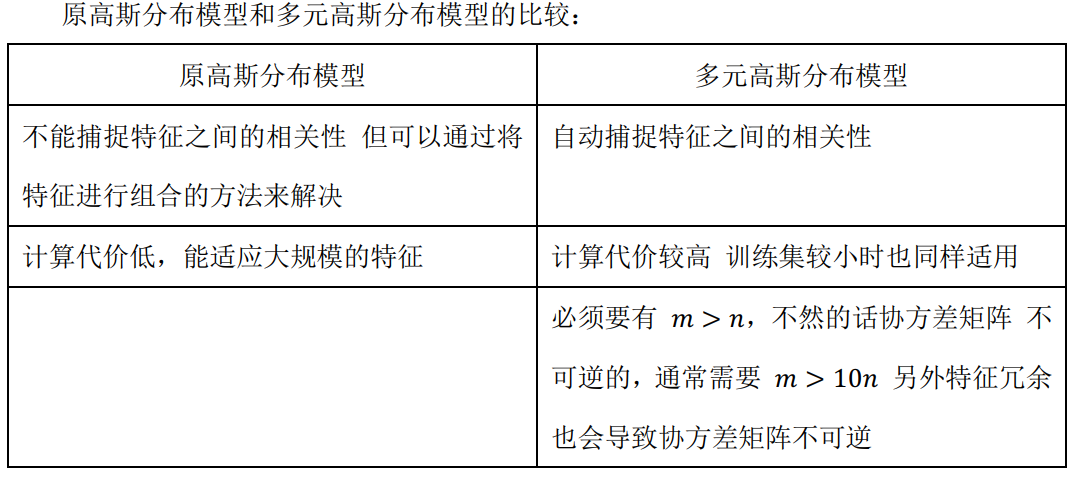
 即：**参数更新**如下所示（吴恩达视频截图）

（2）协同过滤算法

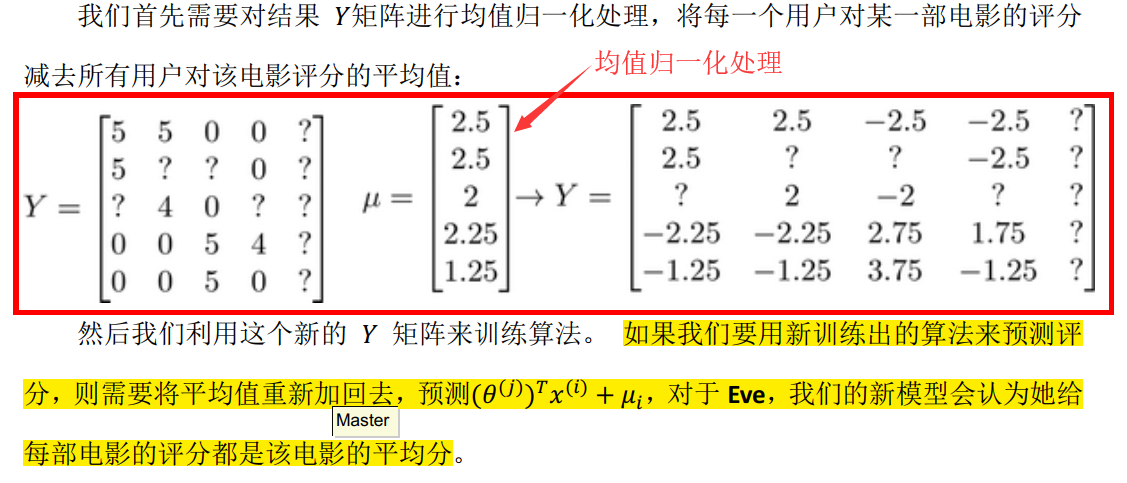


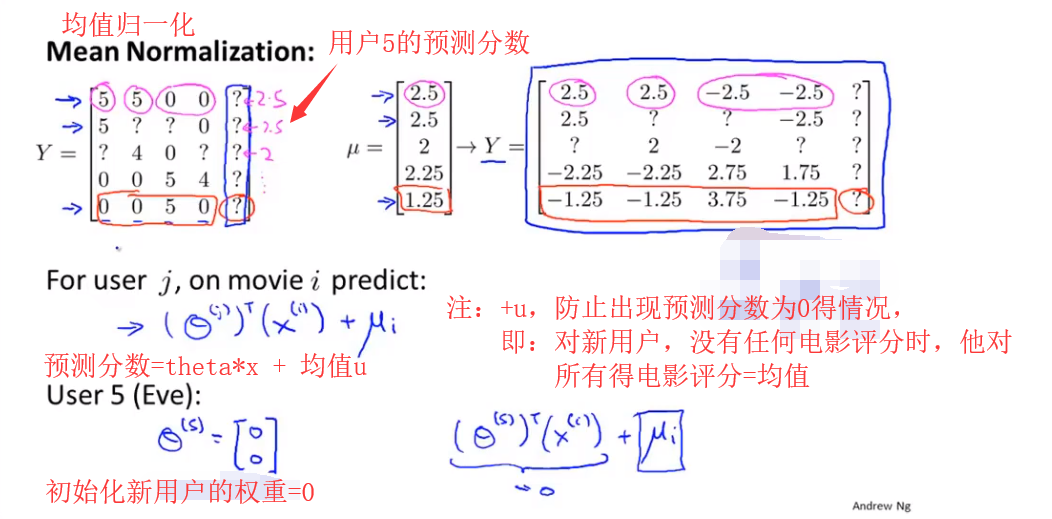
**附：**（参数名：位置、含义）



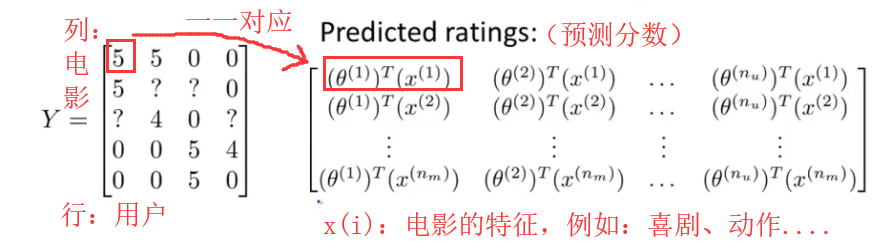


（3）**均值归一化**：**防止出现预测分数=0的情况，使用归一化后的数据进行预测分数**





（4）预测评分：向量化——低秩矩阵分解



（5）SVD在推荐系统中的应用

先使用**SVD**来对数据进行**降维**，然后进行**相似度计算**，提高推荐效果。

关于奇异值个数n的选择：

在进行SVD分解后，求解保留奇异值矩阵90%的能量信息—--即：

先将奇异值求平方，再进行

求：（n个奇异值平方和）/（所有奇异值平方和）≥90%，即得到n。