2. 列举你知道的无约束最优化方法（至少三个），并选一种方法进行详细介绍

分析：梯度下降法，牛顿法，拟牛顿法，共轭梯度法... （介绍略）

3. 请简要叙述正则化项中的 L1 和 L2 方法

分析：1 正则化和 L2 正则化可以看做是损失函数的惩罚项。L1 正则化是指权值向量 w 中各

个元素的绝对值之和，通常表示为||w||1。L2 正则化是指权值向量 w 中各个元素的平方和

然后再求平方根（可以看到 Ridge 回归的 L2 正则化项有平方符号），通常表示为||w||2。

L1 正则化可以产生稀疏权值矩阵，即产生一个稀疏模型，可以用于特征选择

L2 正则化可以防止模型过拟合；一定程度上，L1 也可以防止过拟合

4. 简述 k-means 的主要优缺点及针对缺点的优化方案

分析：优点：经典、简单、快速、对密集簇效果较好

缺点：对 K 值敏感，且只适用于能求距离均值的应用，不适合非凸簇或大小差别很大的簇

改进：二分 k-means，k-means++...

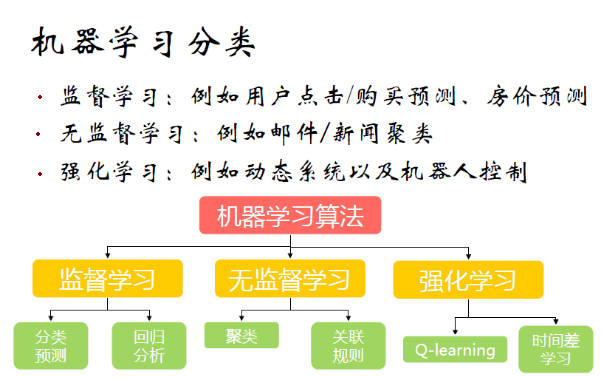
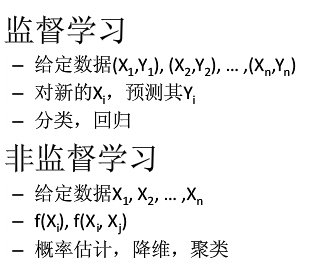
5. 简述 UserCF 的主要步骤

分析：一、找到和目标用户兴趣相似的用户集合——计算两个用户的兴趣相似度

二、找到这个集合中的用户喜欢的，且目标用户没有听说过的物品推荐给目标用户——找出

物品推荐

6.**有监督学习是从标签化训练数据集中推断出函数的[机器学习](http://baike.baidu.com/view/7956.htm" \t "http://baike.baidu.com/_blank)任务。**

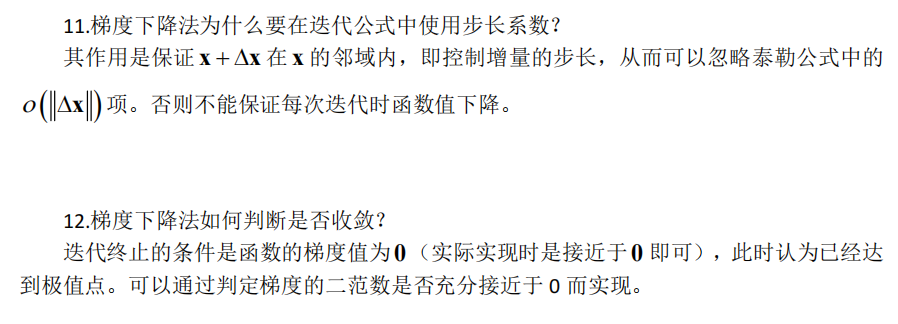


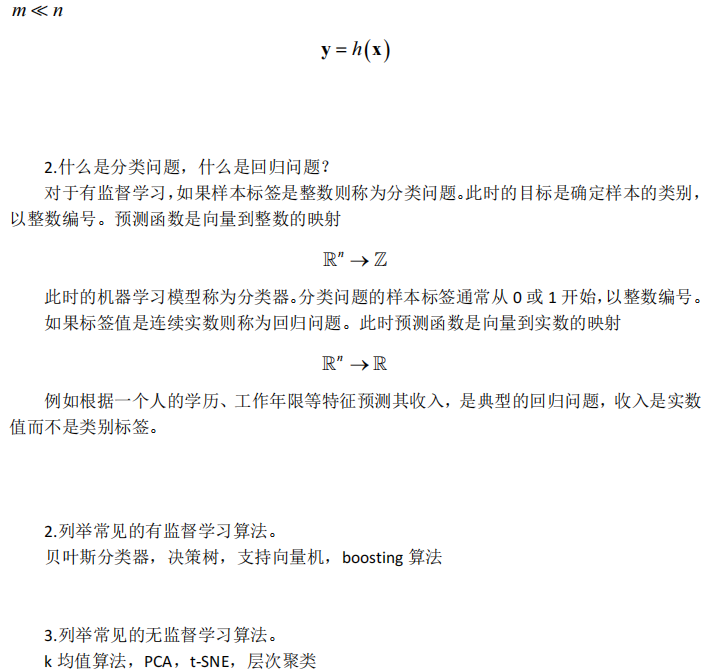
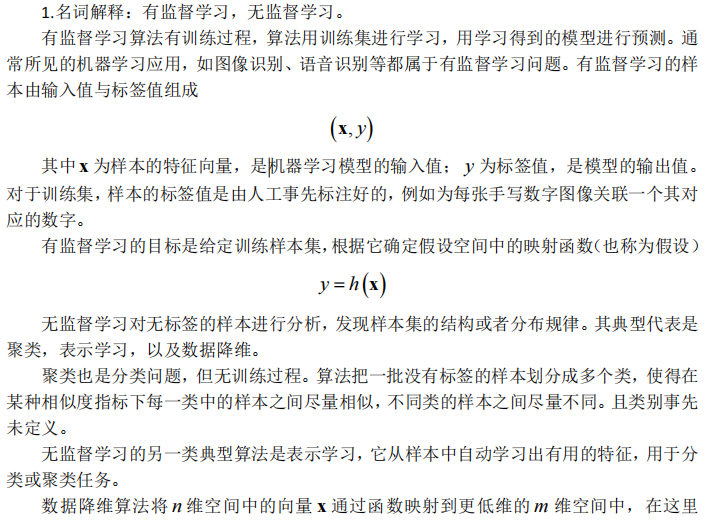
1. **L0范数是指向量中非0的元素的个数。**

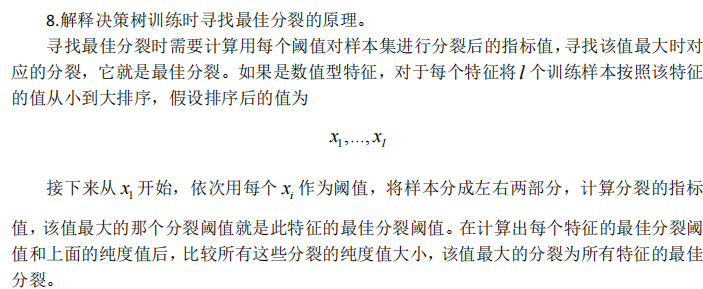
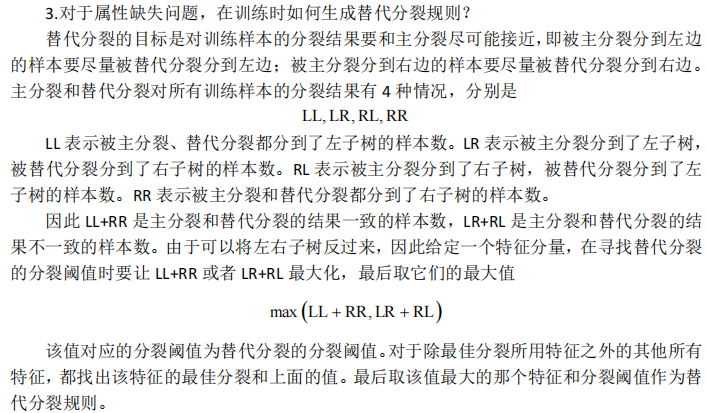
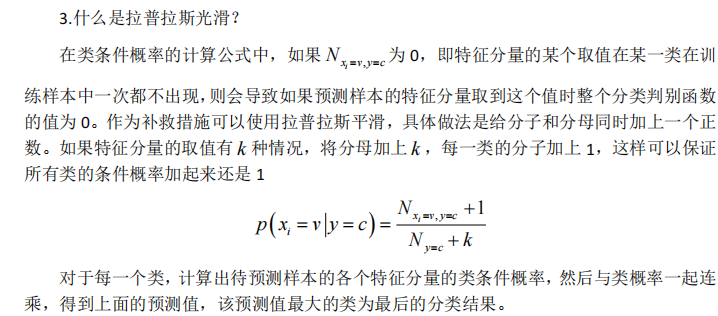
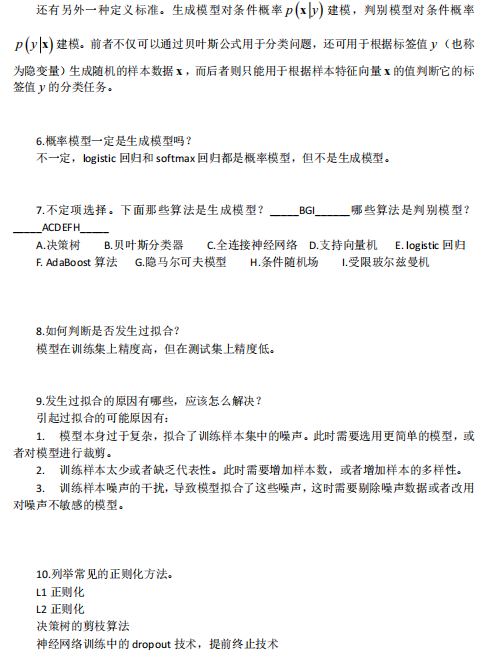
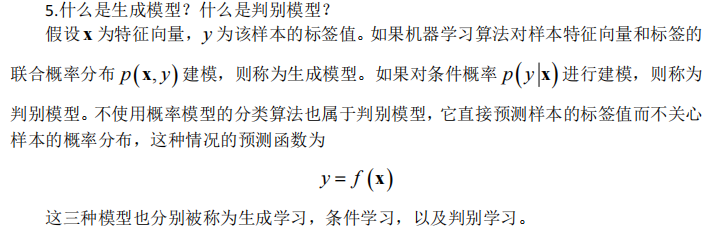
**L1范数是指向量中各个元素绝对值之和，也有个美称叫“稀疏规则算子”（Lasso regularization）。**

**L2范数是指向量各元素的平方和然后求平方根**

1. **交叉验证可以用于防止模型过于复杂而引起的过拟合**

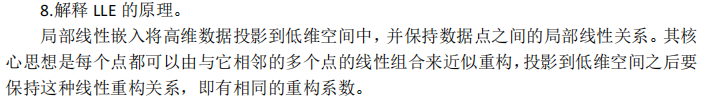
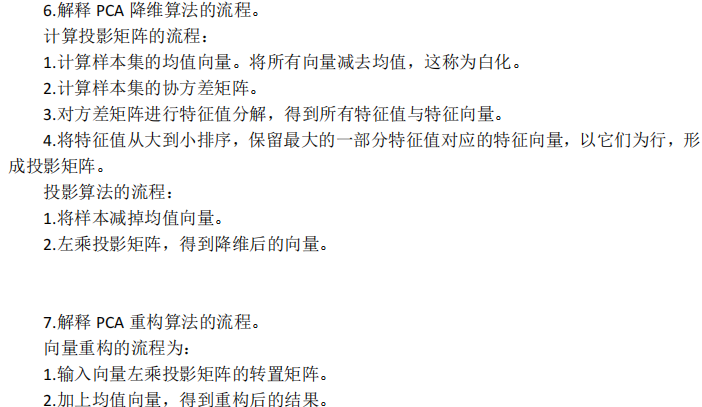




2.简述 k 的取值对 k 近邻算法的影响。

如果其值太小，则容易受到噪声的影响，导致泛函性能下降，出现过拟合。如果 *k* 值等

于训练样数，则对于任意的预测样本，都会将其预测为训练样本集中数量最大的类。

11.解释 KPCA 的原理。核主成分分析是核技术与主成分分析相结合的产物，它使用核技术将向量映射到更高维

的空间，然后用 PCA 进行处理。核技术用一个映射（称为核映射）将原始向量变换到另外

一个空间，然后进行处理。

