**1.决策树 (8分)**

(1)什么是互信息

(2)**用互信息（信息增益）选择的目的是什么，能带来什么效果**

选择互信息最大属性，使得切分后各个节点的信息熵平均较小，节点样本纯度较高。  
(3)怎么防止过拟合？

1. 提前结束
2. 剪枝处理，包括预剪枝和后剪枝

**2.贝叶斯决策 (8分)**

(1) 对于样本x，其标签为0/1，若x真实标签为0，错误划分为1的损失为a， 若x真实标签为1，错误划分为0的损失为b， **求最优决策方法**

**最小化risk**

**将X分类为1的风险：**

**R(Y=1|X)= 0\*P(Y=1|X)+b\*P(Y=0|X)= b\*P(Y=0|X)**

**将X分类为0的风险：**

**R(Y=0|X)= a\*P(Y=1|X)+0\*P(Y=0|X)= a\*P(Y=1|X)=a\*(1- P(Y=0|X))**

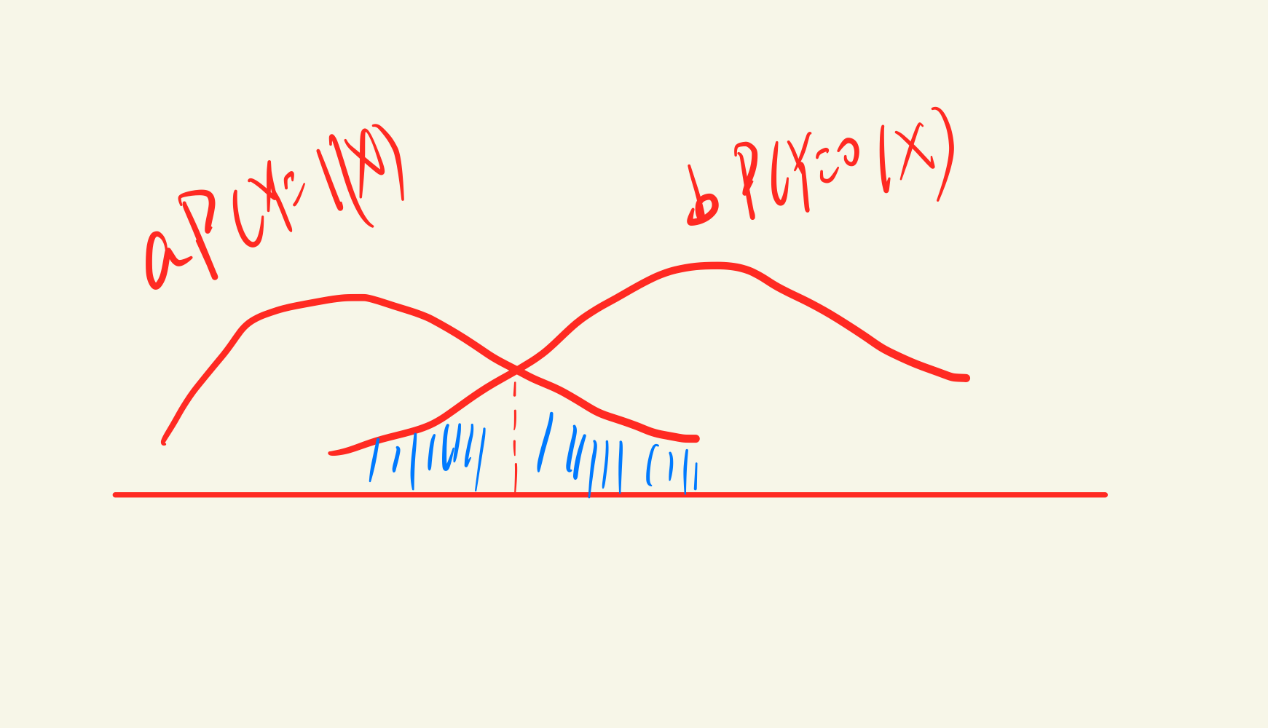
**argmin R(Y|X)= argmin{R(Y=1|X)，R(Y=0|X)}**

**所以分类平面：**

**当h>1 , 即选择分类0**

**当h<=1 , 选择分类1**

(2)结合图画出来。



**3.svm (8分)**

(1)线性可分时的优化目标，约束条件

>1  
(2)线性不可分咋整

(3)**支持向量的意义**

在KTT条件下，只有一部分向量的拉格朗日乘数不为0，这一部分向量叫做支持向量。对于参数W的重构，仅仅依赖于这一小部分支持向量。

可以说，分类器的确定仅仅依赖于这些支持向量。

**4.机器学习的方法 (8分)**

(1)机器学习估计参数模型的方法

MLE：极大似然估计，最大化似然函数P(D|theta)

MAP ：最大后验概率估计,最大化后验概率

P（theta|D）= P(D|theta) P(theta)/P（D）正比于P(D|theta) P(theta)

(2)这些方法有什么不同

MLE将参数假设为一个确定的值，这个值使得出现实验数据的概率最大化。所以需要最大化似然函数，来求解这个参数值。但是当试验数据较少时容易发生过拟合。

MAP将参数看作一个随机变量，引入先验概率，参数值是使得后验概率取最大的随机变量的值。但是当先验概率不一致时，可能有不同参数值的估计。  
(3)**这些方法什么时候趋于一致**

当实验的样本数量N足够多时，先验概率的影响几乎消失，此时MLE和MAP对参数估计的结果趋于一致。

**5.逻辑回归 (7分)**

(1)**逻辑回归为什么优化函数使用条件分布log(Y|x)而不是联合分布log(x,Y)**

**P(X,Y|W)=P(Y|X,W)P(X,W)/P(W), P(X,W)/P(W)的求解过于困难或无法求解。但是P(Y|X,W)是逻辑回归所推导的，容易求解。**  
(2)逻辑回归的推导

1. 各类符合高斯分布
2. 条件独立

**6.pca (8分)**

(1)从**信号重构**角度推导pca

最大可分

最近重构  
(2)怎么实现信号压缩

**7.k-means (8分)**

(1)k-means流程

0.输入样本和划分的类别数K。

1.对于K个类别，随机选择K个样本作为初始类别均值向量

2.对于N个样本，计算其到每个类别均值向量的距离d=||x-u||2

3.将样本分类到距离最近的均值向量所属类别

4.对于K个类别，重新计算均值向量

5.若均值向量没有发生改变，输出分类结果，否则跳到第2步。   
(2)k-means目标函数

误差：  
(3)**k-means和GMM的？em算法的不同**

**K-means是基于非概率模型的分类，分类的结果就是决策函数的结果，属于硬分类（hard assignment）**

**EM算法是基于概率的分类，通过计算一个样本属于某类的概率，选择样本所属类别概率最高的作为分类结果，是软分类（soft assignment）**

**8.线性回归 (5分)**

实验一是非线性回归，**怎么将非线性回归变成线性回归**

使用泰勒展开，将非线性的sin(x)，转化为多项式之和f=W^t\*x

其中x=(x^0,x^1….)