**(a)Give two examples of supervised learning problems and two examples of unsupervised learning problems. Indicate which ones are the inputs and which ones are the outputs in each problem.**

**Supervising**

**1.1** According to the location of houses, size of houses, is there a parking space predict the price of the houses.

**Input:** location of houses, size of houses, is there a parking space

**Output:** the price of the houses

**1.2** According to the Petal size, Petal colour, Petal shape predict the types of irises.

**Input:** Petal size, Petal colour, Petal shape

**Output:** Types of irises.

**Unsupervising**

**2.1** The google news collected many news, divide the news into many clusters according to the theme of the news，the put together the news in the same cluster.

Input: many news

Output: put together the news in the same cluster.

**2.2** Large computer clusters: Finding machines that tend to work together and putting them together will increase efficiency

**Input**: Large computer clusters

**Output**: put together the machines that tend to work together.

**(b)Let *X* and *Y* be two random variables with joint probability distribution *P* (*X; Y* ).Show that the mean of their sum satisfies E[*X* + *Y* ] = E[*X*] + E[*Y* ]**

**Given P(x;y), we can get P(x) and P(y).**

**Continuous variable:**

**贴图**

**F(x)=（2/3 \* 2/3 \* 2/3 \*3/10）/（7/10\*6/10\*6/10）=**

**Prior: ( P(A) )according to the experience, to Predict the probability of something that doesn't happen but will happen.**

**Likelihood( P(B/A)/P(B) ): an adjustment factor that brings the estimated probability closer to the true probability.**

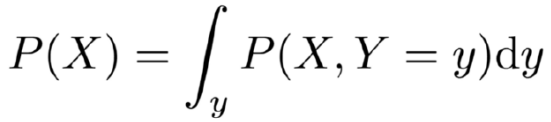
**marginal likelihood: Is a likelihood function that marginalizes certain parameter variables**

**posterior( P(A/B) ): after thing B happen, the probability of thing A happen**

**Maximum likelihood estimation：** a method that determines values for the parameters of a model. The parameter values are found such that they maximise the likelihood that the process described by the model produced the data that were actually observed.

**Marginalisation**： is a method that requires summing over the possible values of one variable to determine the marginal contribution of another.

离散变量：符号改一下就成了

连续变量：

朴素贝叶斯之所以朴素 是因为：Assume All features are independent of each other

http://latex.codecogs.com/gif.latex?%5Clarge%20%24%24%7BP%7D%28%7BC%7D%7C%7BF_1%2C...F_n%7D%29%3D%20%5Cfrac%7BP%28C%29P%28F_1%2C...F_n%7CC%29%7D%7BP%28F_1%2C...F_n%29%7D%24%24 **===**http://latex.codecogs.com/gif.latex?%5Clarge%20%24%24P%28C%7CF_1%2C...F_n%29%3D%5Cfrac%7BP%28C%29%5Cprod_%7Bi%3D1%7D%5E%7Bn%7D%20P%28F_i%7CC%29%7D%7B%5Cprod_%7Bi%3D1%7D%5E%7Bn%7DP%28F_i%29%7D%24%24

**It would be inconvenient and difficult to calculate the maximum value directly.**

**PCA：**

**Maximum variance theory：Select K-dimensional features to maximize the variance of the features in each dimension**

**最小平方差其实就是该数据点投影到转化后坐标轴的距离。让距离最大化就行了（不懂老师问什么）**

**W2 最大的，因为 特征值越大，方差投影越大（我都不懂这题目到底想问什么）**

**3）. 51/（51+6）**

**4）妈的 互不相关， 草 有个什么联系 不懂这老师想问什么**

**5） 归一化，求出协方差和特征值特征向量，按照大小排列， Y =x\*（排列的特征向量）**

**6）他这么写 是要降到几维啊 一维吗？ 一维没有协方差了啊**

**Odds:**

**the probability that a particular outcome is a case divided by the probability that it is a noncase.**

* 1. **？？**

**80% ？？**

**二分类做多分类 比如（有ABC 三类）， A首先 和BC 各一类，就分出了A 类， 其次把 B 和AC 各分一类，拿出来B, 最后把C和 AB各分一类 就拿到了C 二分类就成为了多分类。**