1. 如果概率密度函数满足：

$$p(x_1,x_2,x_3,x_4,x_5) = p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_1,x_3)p(x_5|x_3,x_4)$$

画出对应的贝叶斯网络，马尔可夫随机场和因子图。

2. 如下表数据前四列是天气情况（阴晴 outlook，气温 temperature，湿度 humidity，风 windy）；最后一列是类标签，表示根据天气情况是否出去玩。

outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes

1）根据上述训练数据，基于信息增益决策树应该选哪个属性作为第一个分类属性？

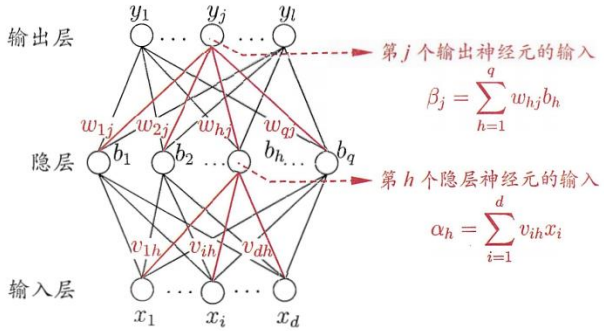
2）请画出两层决策树模型。

3）使用朴素贝叶斯方法预测测试样本（outlook=rainy, temperature=cool, humidity=normal, windy=FALSE）的类标号

3. (神经网络)回答以下问题

(1) 推导下图中 BP 算法中 $\theta_j, v_{ih}, \gamma_h$ 的更新公式。

(2) 试述学习率对神经网络训练的影响。



4. 头条问答是一个新兴的移动社交问答平台，其将信息和人精准匹配，为问题找到合适的回答者；为回答找到合适的阅读者。目前，头条问答每天已有数万用户参与答题，带来的优质回答每天有数千万的阅读。因此一个重要问题就是如何为每个热门问题找到愿意回答的专家用户并将问题推送给他们。

现有数据如下：

a) 用户标签文件，每一行代表一个专家用户，包括四个属性：

i	加密的专家用户 ID：	专家用户唯一标识
ii	专家用户标签	标签包含多个
iii	词 ID 化序列	将专家描述文本删除语气词和标点并分词，再将分词后的每个词用一个 ID 替换。
iv	字符 ID 化序列	将专家描述文本删除语气词和标点并分词，再将分词后的每个字用一个 ID 替换。

b） 问题数据文件，每一行代表一个问题，包含七个属性：

i	加密的问题 ID：	专家用户唯一标识
ii	问题标签：	标签包含多个
iii	词 ID 化序列：	将问题描述文本删除语气词和标点并分词，再将分词后的每个词用一个 ID 替换。
iv	字符 ID 化序列：	将问题描述文本删除语气词和标点并分词，再将分词后的每个字用一个 ID 替换。
	点赞数：	问题所有答案的点赞总数，可以表明问题的热门程度。
	回答数：	问题最终有多少个回答，可以表明问题的热门程度。
	精品回答数：	问题最终有多少个精品回答，可以表明问题的热门程度。

c） 问题分发数据，每一行代表一条问题推送记录，由一条加密的问题 ID，一条加密的专家 ID 和专家是否回答的标记（0 表示忽略，1 表示回答）组成。

问题如下：

(1) 对专家的问题推荐，可以使用哪些机器学习算法？

(2) 你认为哪些数据对推荐结果有较大影响？

(3) 给出你认为合理的方案，言之有理即可。