

6.034 期末考试

2002 年 秋

姓名	
E-mail	

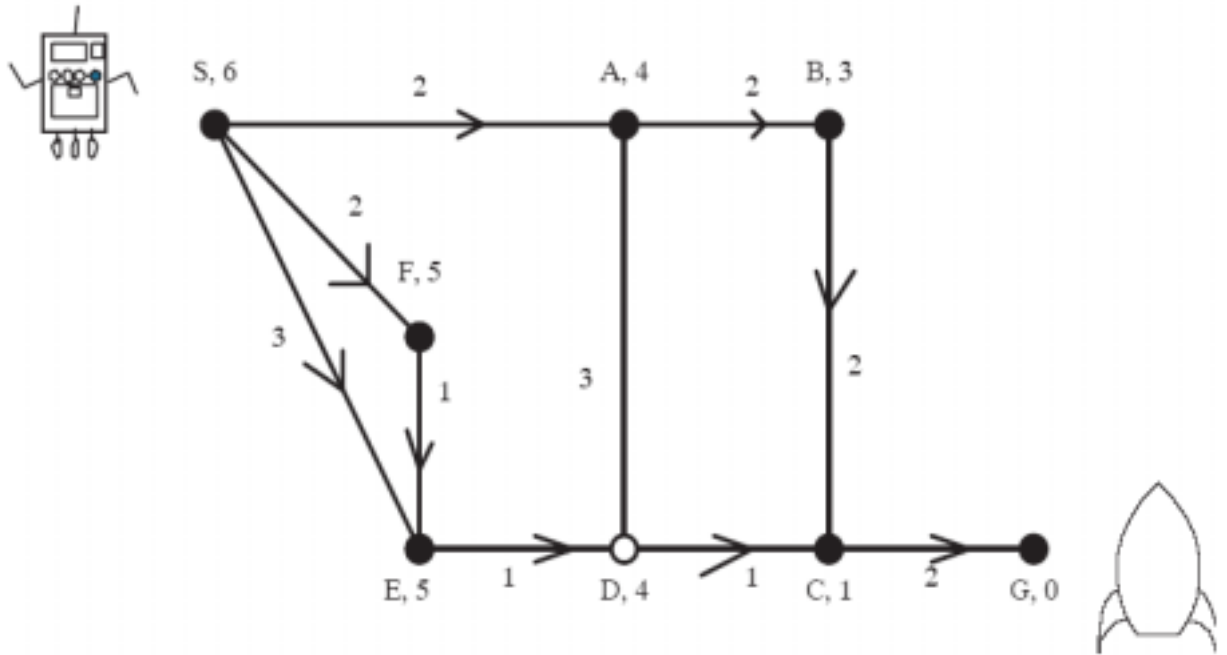
注意：每个问题都有一些简单的部分，但有一些问题的结束部分比较有挑战性；所以你应该尽量保证把简单的做完。

问题号	所占分数	得分	级别
问题 1	16		Phw rcb kk jb sl js as ou ml
问题 2	22		Phw rcb kk jb sl js as ou ml
问题 3	14		Phw rcb kk jb sl js as ou ml
问题 4	21		Phw rcb kk jb sl js as ou ml
问题 5	13		Phw rcb kk jb sl js as ou ml

问题 6	14		Phw rcb kk jb sl js as ou ml
总分	100		

问题 1：搜索（16 分）

Wallace 是一个机器人，他完成了自己在月球上的旅程，将要乘坐太空飞船（在图中的 G 点下方）启程回地球了。Wallace 必须迅速的从 S 点到太空飞船所在的 G 点。他要通过地上的标记来探路。在 D 点有一个会闪的交通灯，在这个问题里，D 点只和 G 点有关系。



注意：这张地图不是按比例画的，所有的线路都是单向的，A-D 的那条除外。各个端点之间的距离用连线旁边的数字标出，节点旁边的数字表示预计剩余的路程。

对于每种搜索方法都做如下的假定：

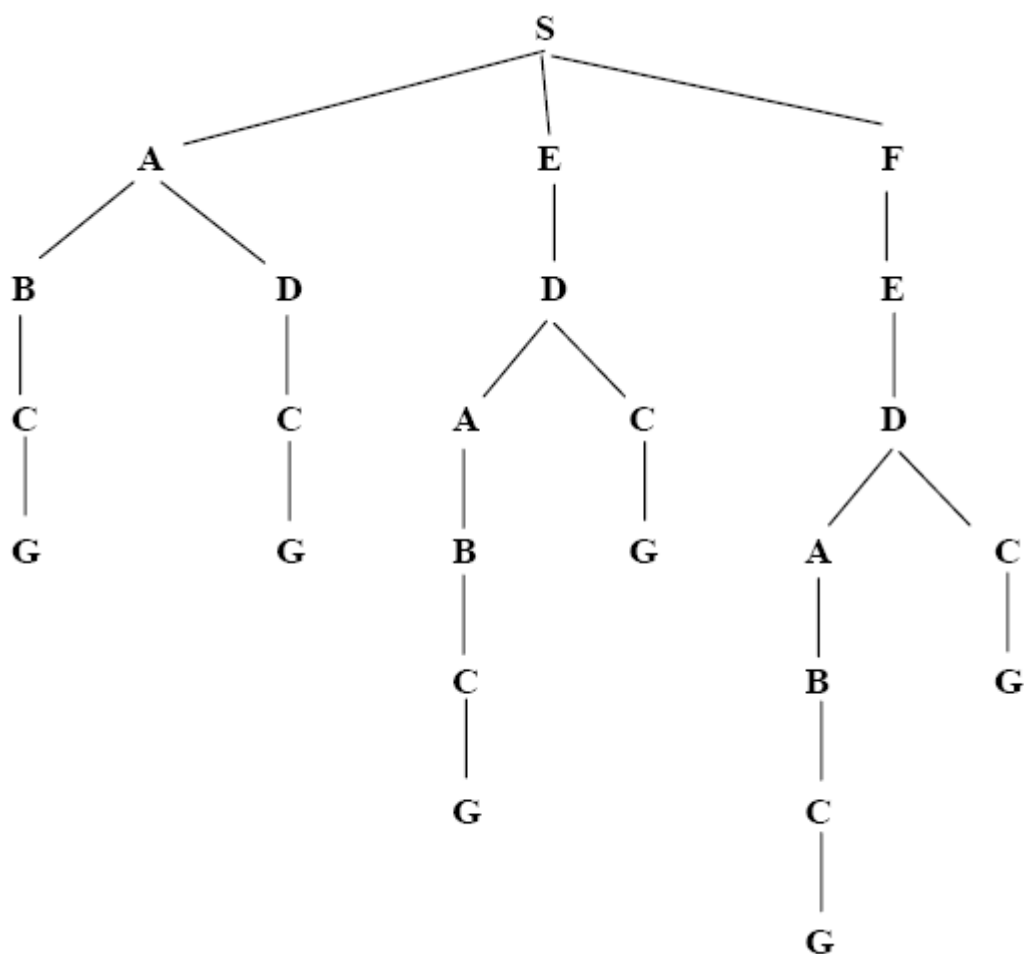
- 不管什么搜索方法都不会产生循环路径；
- 每当向前延伸一段新的路径的时候，就自动检查是不是已经到达终点，如果到达，就停止搜索。其他的时候不会检查是不是到达终点。

下面是解题的一般规则：

- 首先，按照字母的先后顺序，；从第一个节点开始，列出每一个节点的下一个节点。
- 然后，从头到尾的排好队。

A 部分(2 分)

画出完整的，非循环的搜索树。



B 部分 (2 分)

如果采用深度优先的原则进行路径的搜索，节点延伸的顺序会是什么样子的？允许进行备份，但是不提供队列列表和扩展列表；采用这种方法后，它将走一个什么样的路线？

延伸的节点：S A B C G

路径：S A B C G

C 部分 (2 分)

如果采用广度优先的原则进行路径的搜索，节点延伸的顺序会是什么样子的？这里允许提供一个队列列表；采用这种方法后，它将走一个什么样的路线？

延伸的节点 : S A E F B D C G

路径: S A B C G

D 部分 (2分)

使用分支限界法来寻找最短路径 (不使用节点旁边的表示预计剩余路程的数字), 使用扩展列表, 节点延伸的顺序? 选取什么样的路径? 这里假定新路径加在队列的前面。

延伸的节点 : S A E F B D C G

路径: S F E D C G

E 部分 (2分)

使用 A^* 搜索法来寻找最短路径 (允许使用节点旁边的表示剩余路程的数字), 使用扩展列表, 节点延伸的顺序? 选取什么样的路径? 这里假定新路径加在队列的前面。

延伸的节点 : S A B C F E D G

路径: S A B C G

采用 A^* 方法搜索得到的路径的长度是多少?

8

F 部分 (2分)

从 S 点到 G 点的最短路径是什么?

SEDCG or SFEDCG

长度是多少?

7

G 部分 (4 分)

机器人发现, 用 A^* 方法找到的是最快的路径, 但是不是最短的路径。因为它知道在闪光下等待的时间可以走两个单元长度的距离了。或许, 从路径长度的角度来看, 节点旁边标识的关于剩余路径的数字是不可接受的, 但是如果从怎样才能最快到达目的地的角度来看, 这个数字还是能够接受的。

下面哪个节点旁边标识的剩余距离从路径长度来考虑是正确的, 在正确答案上画圈。在别的答案上画 X。

A

B

- E
- 没有这样的节点
- 所有的节点

下面哪个节点旁边标识的剩余距离从旅途的时间角度来考虑是正确的, 在正确答案上画圈。在别的答案上画 X。

A

B

E

没有这样的节点

所有的节点

问题 2：约束传播 (22 分)

6. 034 课程结束了，但是你会发现你的感觉被永久地扭曲了。不管你走到哪里，6.034 学期中的一些东西总是在你的眼前浮现。其中就有这么一个例子，几个星期以前，在 MIT 的一个舞蹈社团举办的一个晚会上。。。。。

(倒叙) 演出的时候，你发现在节目与节目的中间会插入一些幽默短剧以便那些跳舞的女孩能够有时间换衣服。这时你轻轻地对自己说，“其实他们不必要这么做，他们只要为节目编排一个合适的顺序就可以了，这就是所谓的约束传播”，你马上四周看看，深怕别人听到你说的话，还好，他们都在认真地看着幽默短剧，根本就没听到你在说什么。于是你开始着手解决这个排序问题，你针对的是一个最简单的晚会，也就是一幕有六场舞的情况。每场舞的时间段分别标做 1、2、3、4、5、6，每场舞分别标做 A、B、C、D、E、F。

A 部分 纯粹的回溯法 (8 分)

首先，你小声的对自己说，‘每场舞只允许在一个时间段中出现’，于是你马上意识到对于每两个时间段来说，这都是一个二值约束的问题。

约束条件 1	一场舞不会在多于一个时间段中出现
--------	------------------

其次，在第一个和最后一个时间段的表演必须是特别精彩的。你意识到这对于时间段 1 和 6 来说是个一元约束。

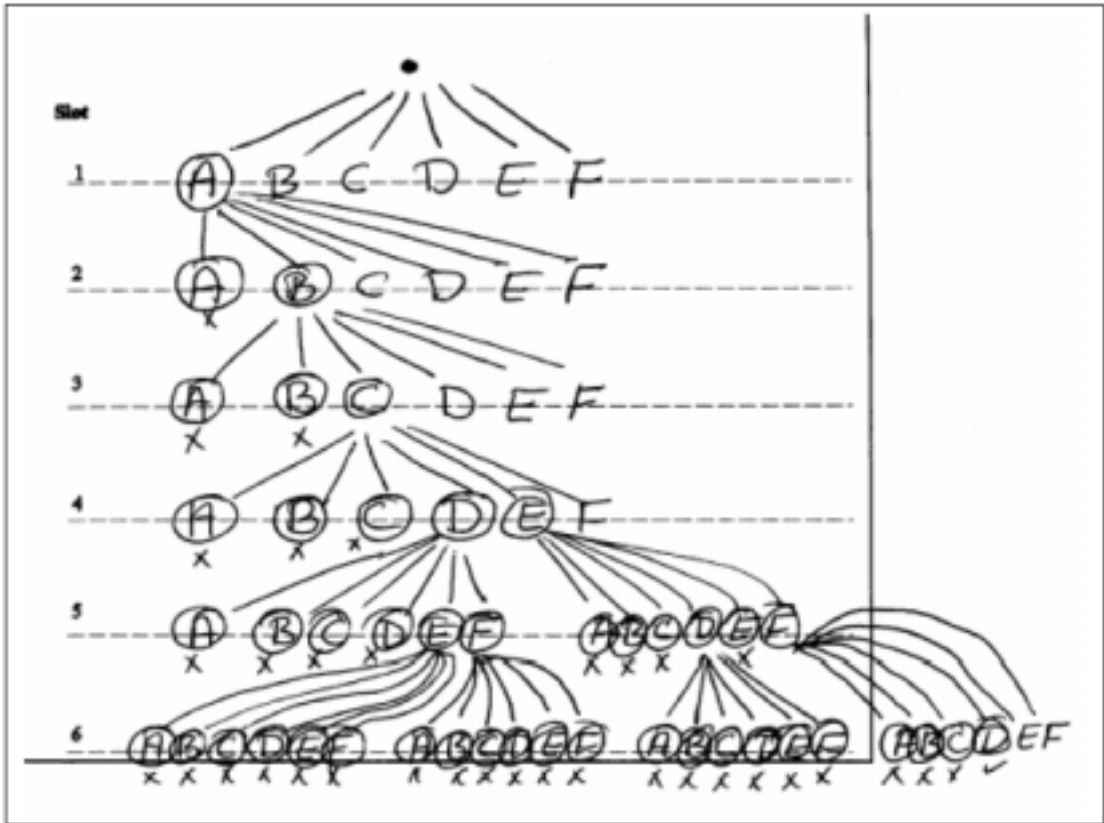
约束条件 2	在时间段 1 和 6 中的舞蹈要特别精彩
--------	----------------------

现在，你已经准备好了为这些舞蹈排一个顺序了。使用下面的数据，用回溯法排出一个符合上面两个规则的顺序。(一般来说，按照字母或者数字的顺序来为变量和数值排序)。给出搜索树。

舞蹈	是否很精彩
A	Yes
B	
C	Yes
D	Yes
E	
F	

为方便起见，在试卷的最后重印了所有的约束和表格，它们可以撕下来。

搜索树：



解决方案为：

时间段	1	2	3	4	5	6
舞蹈	A	B	C	D	E	F

B 部分 带有正向检查的回溯(8 分)

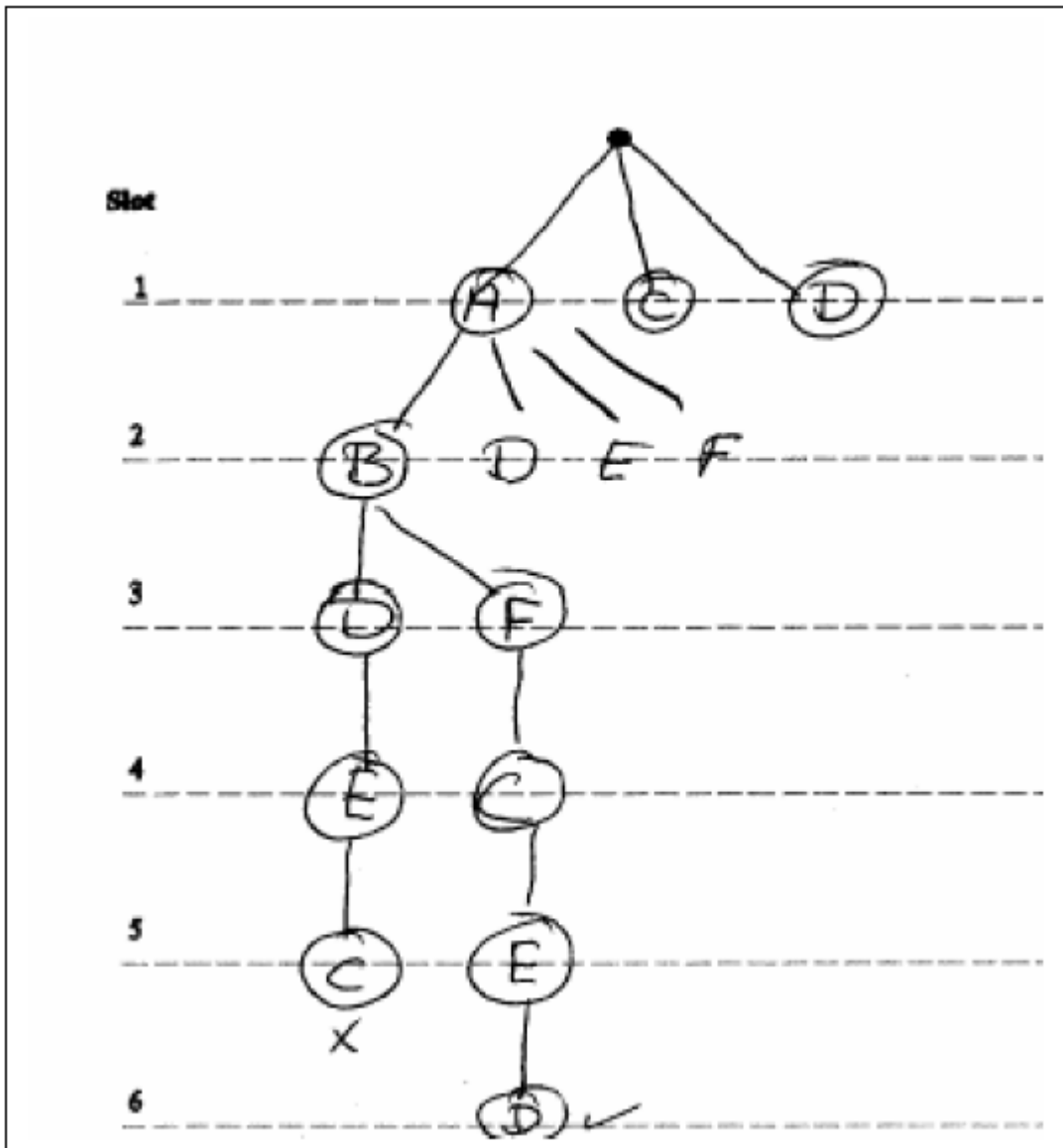
对于取得的成绩，你高兴不已。于是你接着制定规则，你对自己说，‘第三，如果同一个演员参加了两个舞蹈，那么这两个舞蹈不能连在一起’。你意识到这对于两个连续的舞蹈来说是个二元约束问题。

约束条件 3	如果同一个演员参加了两个或两个以上的舞蹈，那么这些舞蹈不能连在一起
--------	-----------------------------------

下面这个表格是关于哪些节目共享了演员的说明：

	A	B	C	D	E	F
A			Shares			
B			Shares		Shares	
C	Shares	Shares		Shares		
D			Shares			Shares
E		Shares				Shares
F				Shares	Shares	

用上面的三个约束条件，通过进行带有正向检查的回溯来给出一个有效的节目安排，给出你的搜索树。



解决方案是：

时间段	1	2	3	4	5	6
舞蹈	A	B	F	C	E	D

C 部分 优化的约束传播 (6 分)

警告： part C 可能会很长，或者可能很难，或者可能既很长又很难；所以你可以考虑先做后面的问题。

现在，你已经达到了你最初的目的了，你欢欣不已，于是你鼓掌叫好，还好，这时候观众也在为幽默短剧鼓掌叫好。Koile 博士是一个舞蹈狂热者，他刚好就坐在你后面，就在下一个舞蹈要开始的时候，她靠上前来，指着你的笔记小声的说，‘如果能够使不同流派的舞蹈交叉进行就更好了’，于是你看看自己已经编排好的这个顺序，真是的，确实是没有注意到这个问题。

这时候，你的助教走上抬来，你决心为按照美学思想找到一个最佳的排列顺序，你对自己说 ‘要是能有一个不同流派的舞蹈之间切换的启发式规则就好了’。于是 Koile 博士提供给你如下的数据：

类型转换的扣分表:

	芭蕾	hip hop	爵士	摇摆舞
芭蕾	10	5	1	2
Hip Hop	5	10	1	2
爵士	1	1	10	3
摇摆舞	2	2	3	10

舞蹈的类型

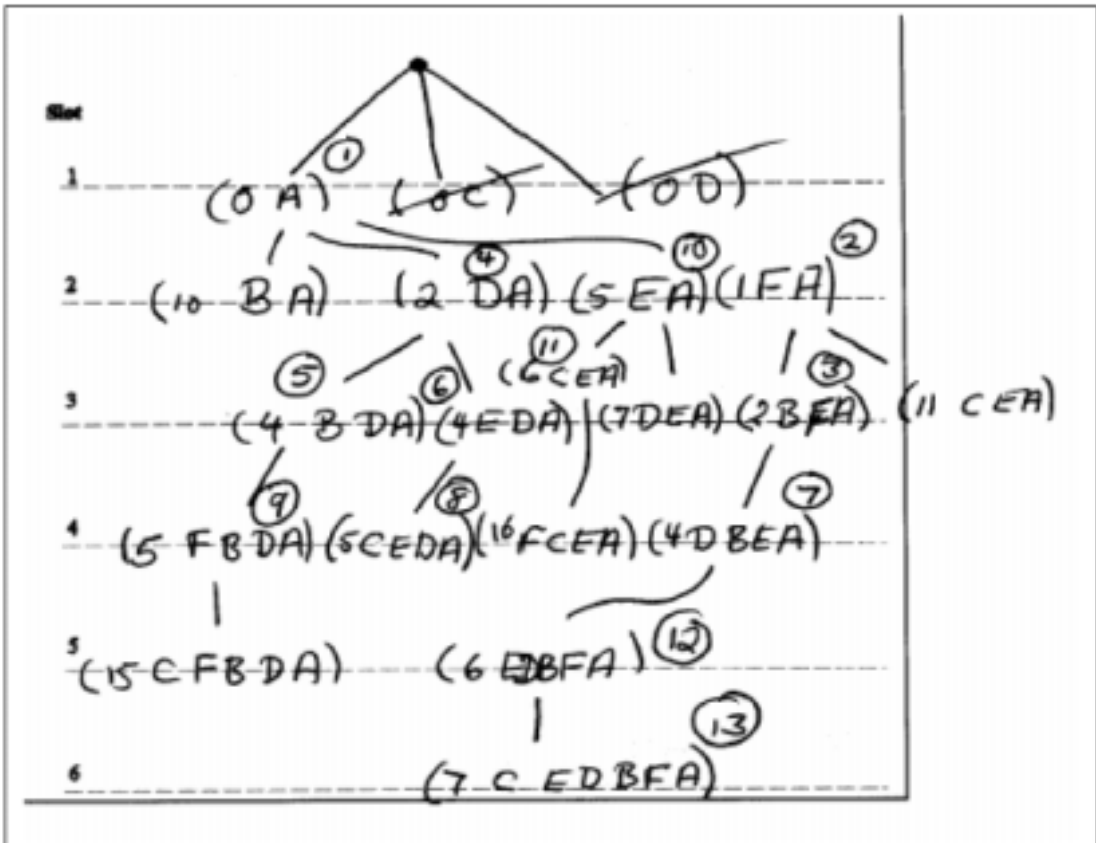
Dance	A	B	C	D	E	F
Genre	Hip Hop	Hip Hop	爵士	摇摆舞	芭蕾舞	爵士

另外, Koile 还对你说, Justin 参与舞蹈 A, 这意味着舞蹈 A 必须排在第一场。

使用上面的这两个表格的数据, 另外舞蹈 A 必须排在第一位, 当然还有前面的三个约束, 试着找出一个最佳的演出顺序 (也就是要被扣最少的分)。给出你的搜索树。

提示: 可以考虑用别的搜索算法, 而不仅仅是深度优先的搜索方法。

在下面的方框中画出你的搜索树 (试卷后面重印了此部分, 并可以撕下):
搜索树



解决方案

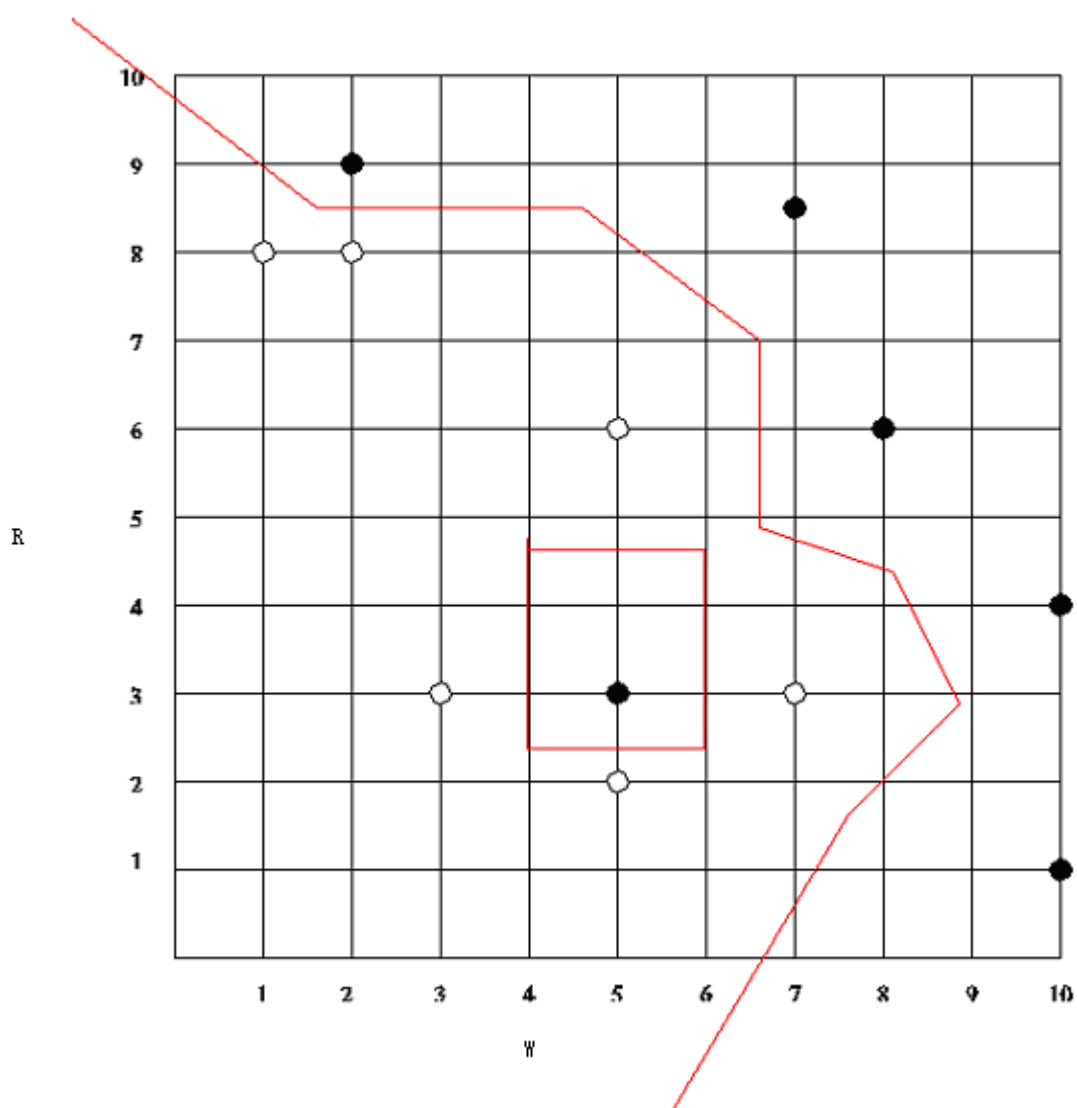
时间段	1	2	3	4	5	6
舞蹈	A	F	B	D	E	C

问题 3：分类（14 分）

A 部分 最近邻（6 分）

6. 034 的教工们打算开办一个电视节目以评选 AI 超级明星，通过这个节目，所有有积极向上的学生都有希望成为 MIT 学生们的偶像。教工们制定了两项评判标准以选择成功的候选人：工作的勤奋与否（W）、智利的高低（R）；教工们将用最近邻分类器把候选者分为 **有潜力成为超级明星**（黑点）和 **一般学生**（空心圆）。

在下图画出通过 1 最近邻分类器在 R-W 图上找到的决策边界。



B 部分 判定树（4 分）

B1（2 分）

现在，先暂时不考虑最近邻。你决定先试试判定树方法。下面，你可以对图示的数据进行两个初始测试，计算出每个测试的平均混乱度。在你的表达式中，允许有 \log_2 这样的，但是不允许有变量。

Test A: $R > 5$:

$$\left(\left(\frac{1}{2} \left(2 \left(- \left(\frac{1}{2} \log \frac{1}{2} \right) \right) \right) \right) \right) \left(\frac{1}{2} \left(2 \left(- \left(\frac{1}{2} \log \frac{1}{2} \right) \right) \right) \right) = 1$$

Test B: $W > 6$:

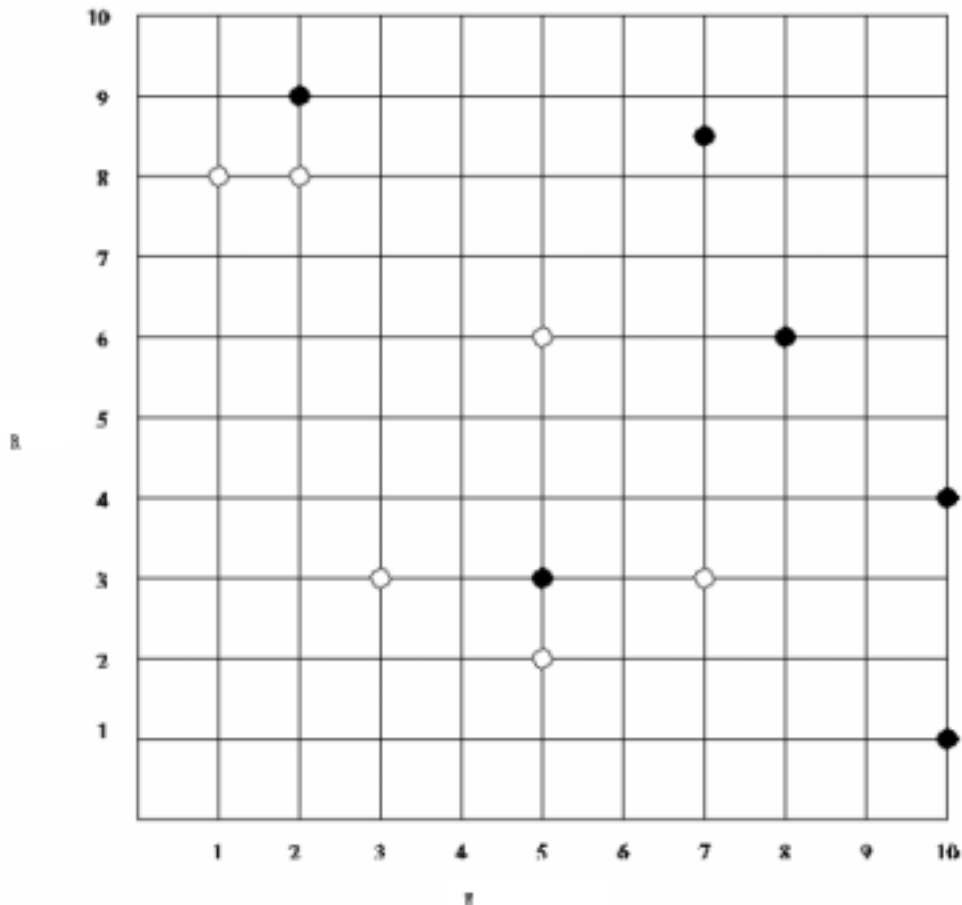
$$\left(\frac{5}{12} \left(\frac{1}{5} \log \frac{1}{5} \right) \left(\frac{4}{5} \log \frac{4}{5} \right) \right) \left(\frac{7}{12} \left(\frac{2}{7} \log \frac{2}{7} \right) \left(\frac{5}{7} \log \frac{5}{7} \right) \right) < 1$$

B2（2 分）

如果采用贪婪算法的话，上面的哪个将首先被选择来构造判定树。

B

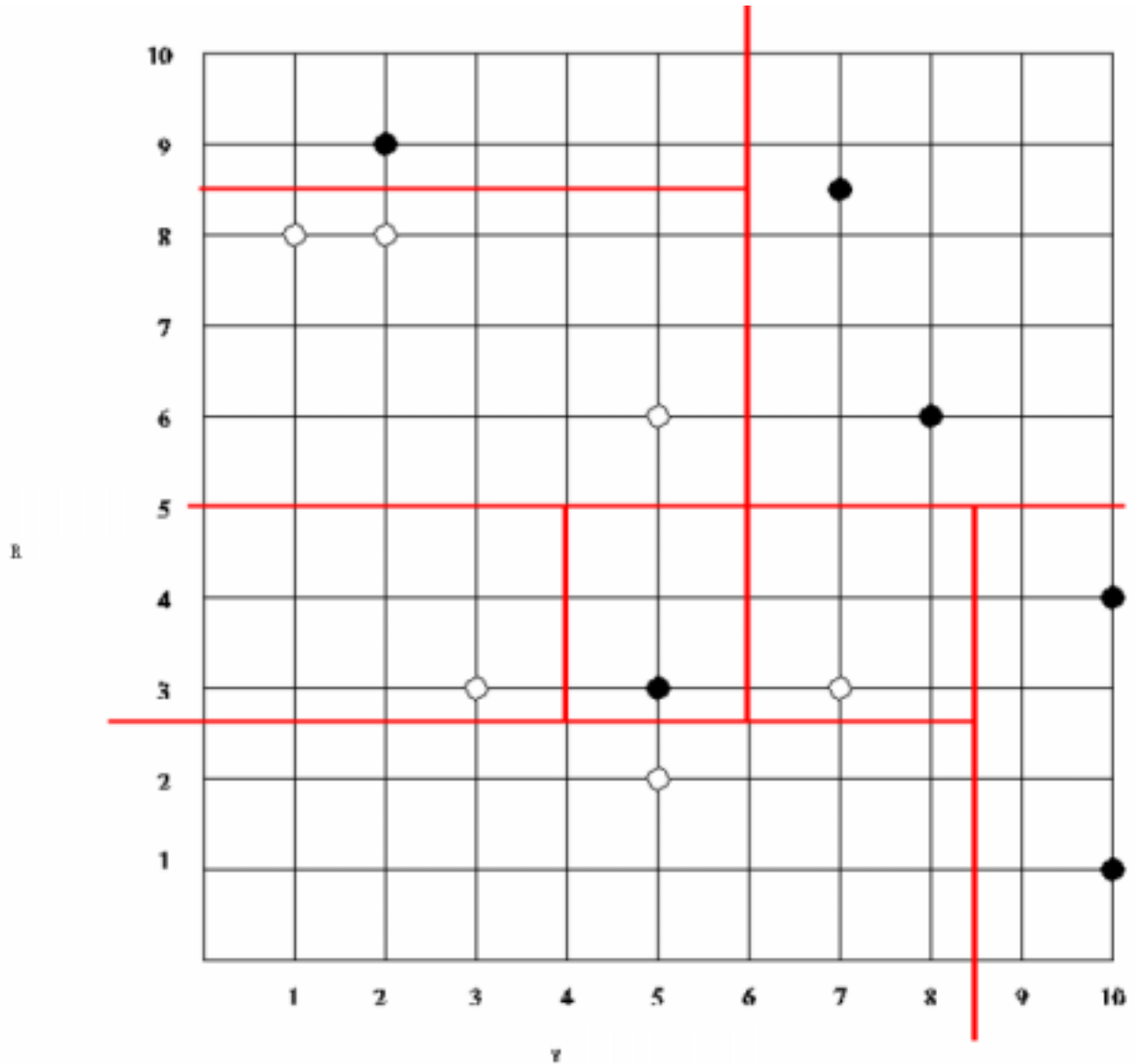
下图是提供给你在上面画草图用的。



C 部分 判定树 (4 分)

现在, 假设判定树构造器选择了 $R > 5$ 作为第一个测试(它可能是对的也可能是错的)。画出判定树构造器将得到 (正确地) 的其他决策边界。

注意: 在分割该图的底端的时候, 别的方法也能够达到同样的效果;



问题 4：神经网络（21 分）

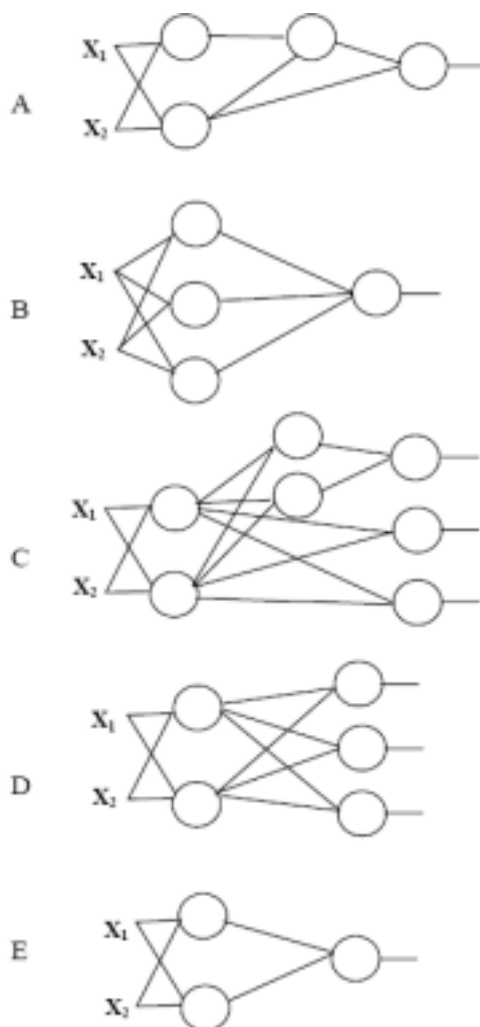
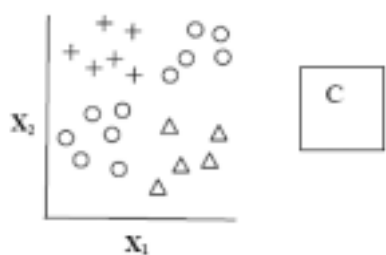
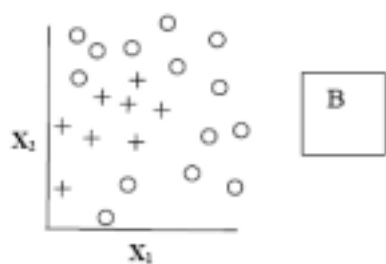
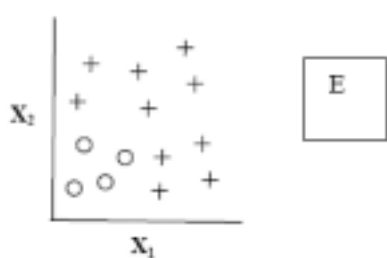
A 部分：感知器（11 分）

A1（3 分）

对于下面的每组数据集，给出使用感知器网络能够将他们完全分类的决策边界的最小数目。

A2（3 分）

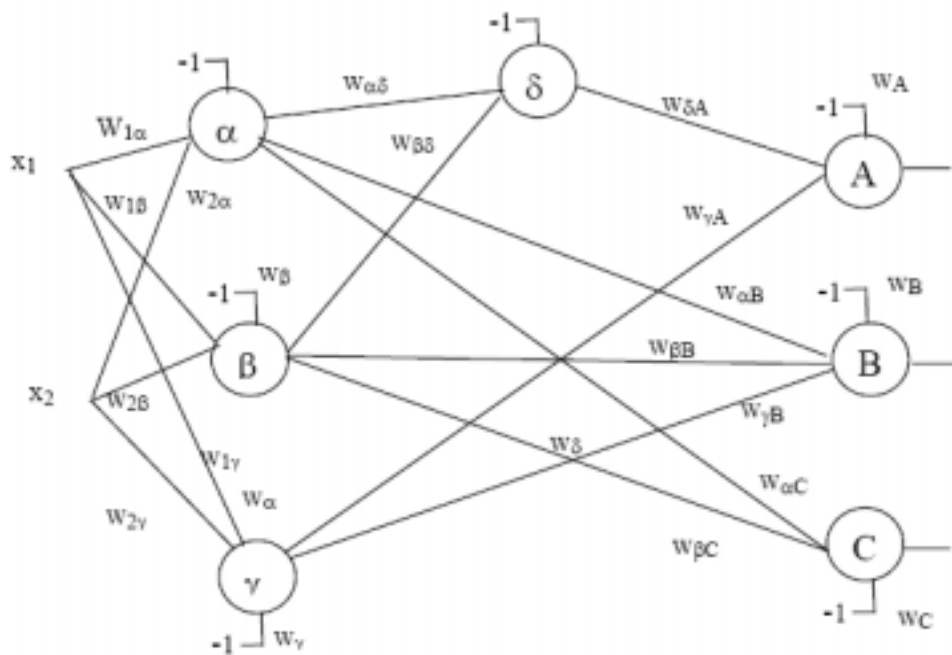
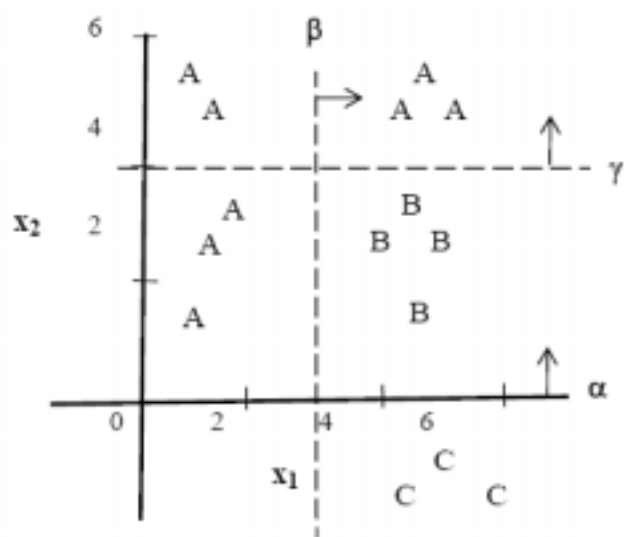
回想一下，感知器的输出是 0 还是 1。下面有三个数据集，对每个数据集都选择一个感知器网络，要求在能够分辨数据集的类别的情况下感知器要最少。在方框里头写上对应的字母。每个网络允许用多次。



A3 (5分)

为下页的感知器网络的节点添上权值。假设：

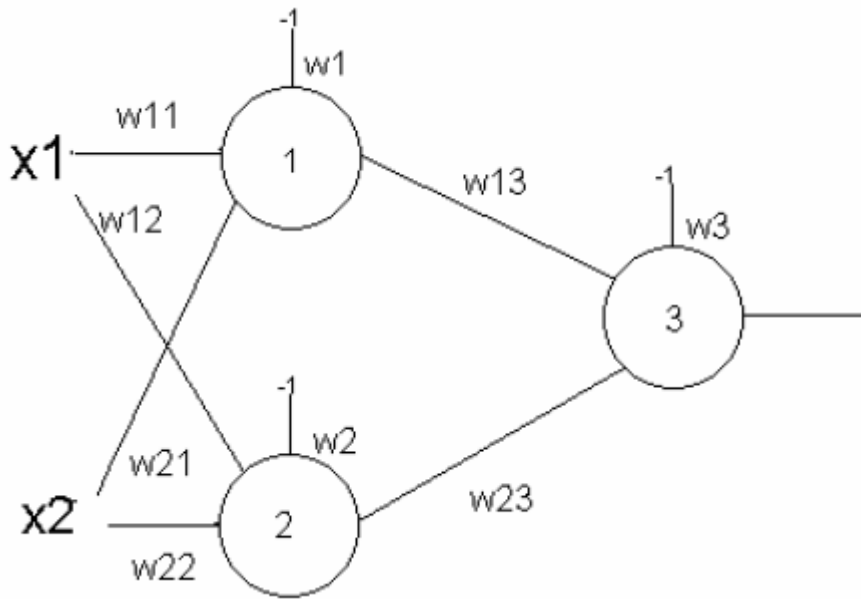
- 感知器的输出为 0 或 1；
- A, B 和 C 是不同的类别
- 那些标有 α (也就是 x 轴), β , γ 的直线表示的是决策边界；
- 图上的那些箭头的方向表示的是边界的方向, 在这个方向上将使得感知器的输出为 1。



α		β		γ		δ		A		B		C	
$w_{1\alpha}$	0	$w_{1\beta}$	1	$w_{1\gamma}$	0	$w_{\alpha\delta}$	1	$w_{\delta A}$	1	$w_{\alpha B}$	1	$w_{\alpha C}$	-1
$w_{2\alpha}$	1	$w_{2\beta}$	0	$w_{2\gamma}$	1	$w_{\beta\delta}$	-1	$w_{\gamma A}$	1	$w_{\beta B}$	1	$w_{\beta C}$	1
w_{α}	0	w_{β}	3	w_{γ}	4	w_{δ}	0.5	w_A	0.5	$w_{\gamma B}$	-1	w_C	0.5
										w_B	1.5		

B 部分：负 s 型单元 (10 分)

下面的 s 型网络有三个单元，分别标作 1, 2, 和 3。每个单元都是一个负 s 型单元，这就是说他们的输出是用表达式 $n(z) = -\frac{1}{1+e^{-z}}$ 来计算的，这个表达式和一般地 sigmoid 不同的地方就是前面多了一个负号。n(z)的导数是： $\frac{dn(z)}{dz} = n(z)(1+n(z))$ 。另外，该网络还用到了一个非标准的误差计算公式 $E = \frac{1}{2}(2y^* - 2y)^2$



B1: 前向传播 (4 分)

使用下面提供的初始权值和输入矢量 $[x_1, x_2] = [2, 0.5]$ ，计算经过前向传播后的每个神经元的输出。计算的时候要用到的负 sigmoid 值在试卷后面的附录的表格中。

权	w1	w11	w12	w13	w2	w21	w22	w23	w3
数值	0.5	1	1	0.5	0.5	1	1	0.25	0.5

y1 (神经元1的输出)) $n(w_{11}x_1 + w_{21}x_2 - w_1) = n((1)(2) + (1)(.5) - .5) = n(2) = -.88$

y2 (神经元2的输出)) $n(w_{12}x_1 + w_{22}x_2 - w_2) = n((1)(2) + (1)(.5) - .5) = n(2) = -.88$

y3 (神经元3的输出)) $n(w_{13}y_1 + w_{23}y_2 - w_3) = n((.5)(-.88) + (.25)(-.88) - .5) = n(-1.16) = -0.24$

B2 反向传播(6分)

将学习速度设置为 1，预期输出为 0，通过计算节点 2 和 3 的 δ 值来对网络进行反向传播，并且写下下表中选定的权值的新值。假设权值的初始值就是 B1 部分中提供的那些值，神经元的输出值如下：

节点 1 的输出， $y_1=-1.0$

节点 2 的输出， $y_2=-1.0$

节点 3 的输出， $y_3=-0.2$

注意：一些可能要用到的公式会在试卷结尾的附录中提供；

用不含导数的表达式表示 δ_2 和 δ_3

$$\delta_3 \quad 4(y_3 - y^*)(y_3)(1 + y_3) = 4(-0.2)(-0.2)(1 - 0.2) = 0.13$$

$$\delta_2 \quad y_2(1 + y_2)(w_{23}\delta_3) = 0$$

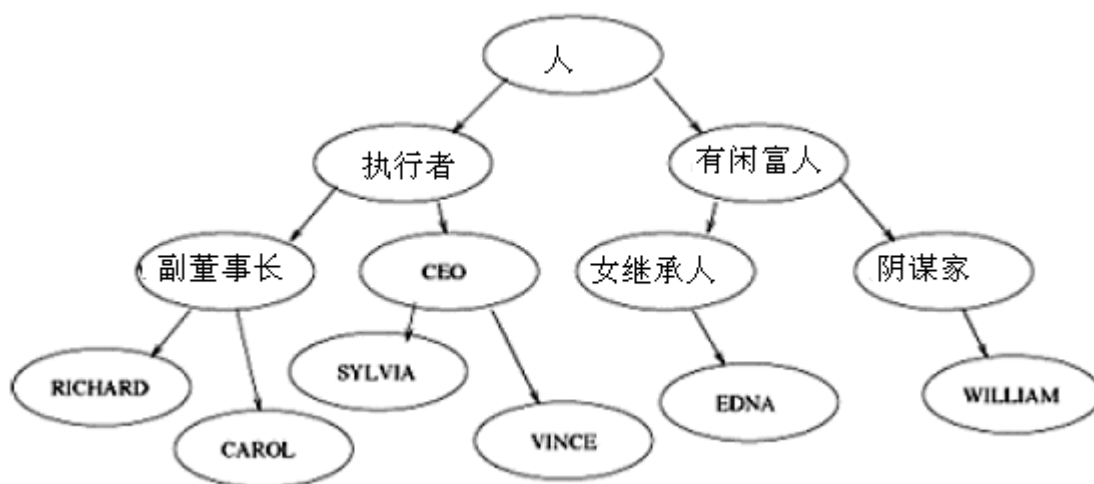
用 δ_s 和数字来表示权值：

权	w_{22}	w_{23}
值	$1 - 0.5\delta_2 = 1$	$0.5 + \delta_3 = 0.63$

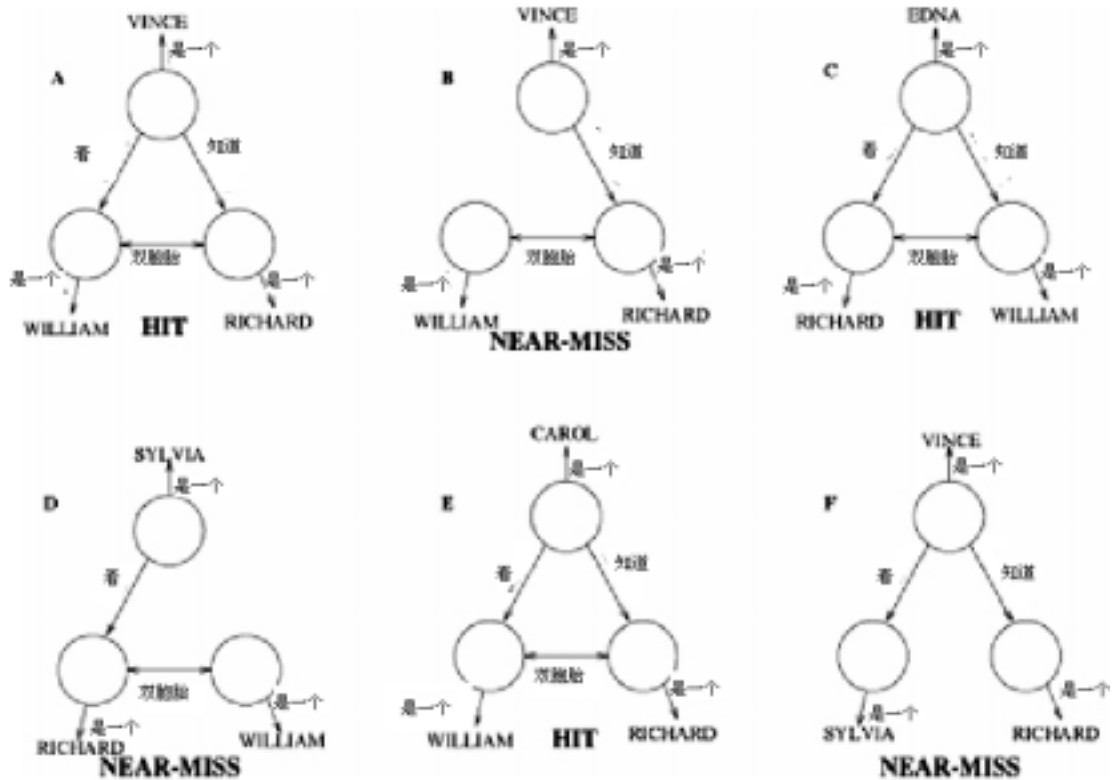
问题 5：似是而非 (13 分)

A 部分：搞错身份 (9 分)

Ben Bitdiddle 是一名企业 AI 工程师，他认为今后的主流产品将是一个肥皂剧的概括系统。这样人们就不用浪费时间来看这些肥皂剧了。作为系统的一部分，他觉得自己该了解一些关于人际关系的概念，于是他决定借助一个似是而非学习系统来做这件事。Ben 打算学习一种喜剧，这个喜剧是关于某人把双胞胎中的一个误认为是另一个后发生的事情。他首先了解在肥皂剧中的人与人之间的关系，并且认为人都是可以归类到一些可预知的类型中。于是他建立了如下的一个用于描述人和人的职业的树。



Ben 用六个**搞错身份**的样本对这个似是而非学习系统进行训练，这些样本已经被我们转化为网络，并且按字母顺序排好了。如下页的图所示：



A1 (6分)

对于上面的每一个样本，Ben 的系统能从已有的数据中构造出什么样的模型，请你画出来。

- 在 A 之后，初始模型建立起来
- 在 B 之后，看 变成 一定看
- 在 C 之后，是 Vince 变成 一定是个人
- 在 D 之后，知道 变成 一定知道
- 在 E 之后，没有变化
- 在 F 之后，双胞胎 变成 一定是双胞胎

A2(3分)

对六个样本指出系统是专有的 (S)、通用的 (G) 或者两者都是 (B)，或者两者都不是 (N)。

A

B.

C.

D.

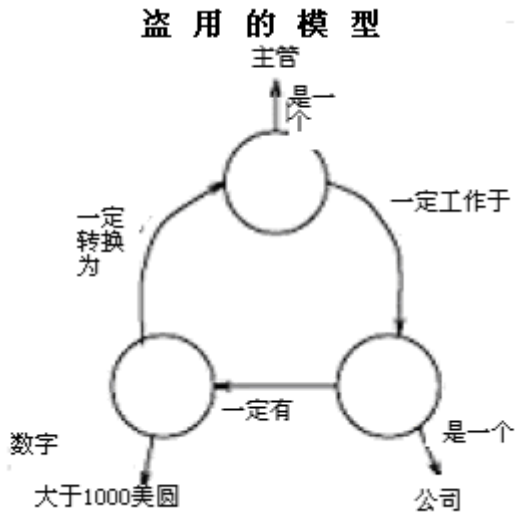
E.

F.

似是而非总是专有的，成功要么什么都不是，要么是通用的。

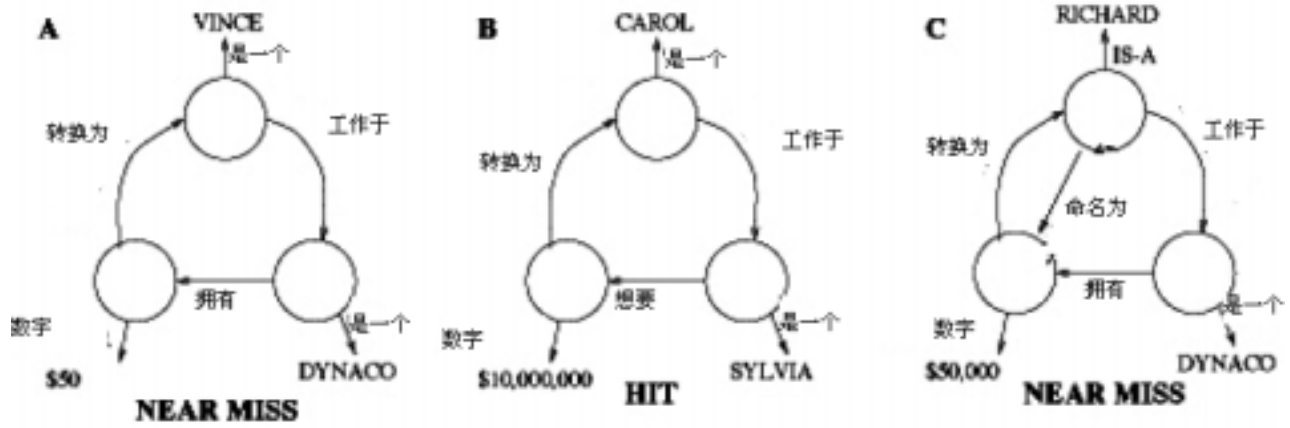
B 部分：盗用（4分）

上一步已经取得了成功，Ben 决定为系统训练另外一个模型。现在 Ben 想做的是当盗用发生的时候，系统能够识别。过了一段时间，他的模型看起来象这样：



B1（2分）

请注意，小额款项的支出常常是为了员工的娱乐花费。那么现在要你决定，Ben 将选择下面的哪个模型作为下一个模型？



B2 (2分)

从你的选择中描绘出新的模型：

将 命名为 改为 一定命名为

对左边那个似是而非，做为盗用来说，50 美元显得太少了，其他没什么差别；

对于中间那个，停止连接会去掉所有关系，会显得比较怪；

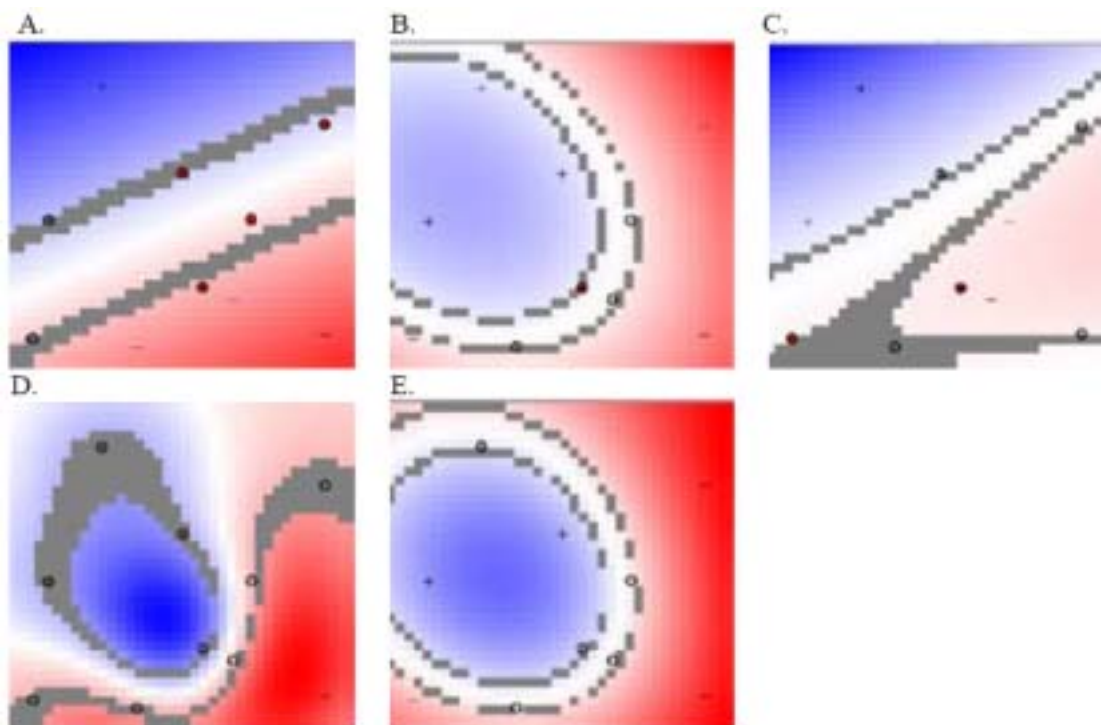
问题 6：支持向量机（14 分）

A 部分：（2 分）

下面的图形表示一些支持向量机，他们将被训练来在同一个数据集当中区分加号（+）和减号（-）。每个图的左下角是原点。对训练数据来说，哪个是最好的分类器？看彩图能够看的更清楚一些。

给出你的选择

E 为每个点给出了适当的值。其实 D 也是，但是 D 有过拟合的危险。



B 部分：（5 分）

为下面的核从 A 部分中选择一个匹配的图：

- 径向基函数, $\sigma = .08$ D
- 径向基函数, $\sigma = .5$ E
- 径向基函数, $\sigma = 2.0$ B
- 线性函数, A
- 二次多项式 C

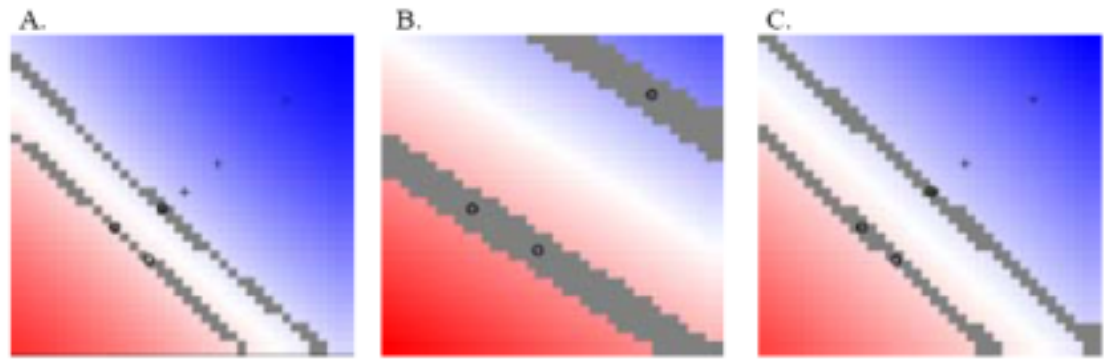
C 部分：（3 分）

按照支持矢量的最小权值到最大权值对下面的图进行排序，假设所有的图都是由相同的机制使用线性核来生成的（也就是说，不是从点积空间转换过来）。

原点还是在每个图的左下角。支持矢量权值又叫做 α_i 值，或者拉格朗日乘法器。看彩图能够看的更清楚一些。

最小 中等 最大

B	C	A
---	---	---



D 部分：（4 分）

假设一个区分加号和减号的支持向量机在点 $x_1 = (1, 0)$ 找到一个加号支持向量，在点 $x_2 = (0, 1)$ 找到一个减号支持向量。

现在要你确定分类向量 w 和阈值 b 的值。在你的关于 w 的表达式中允许包含 x_1 和 x_2 ，因为他们是已知的。但是不允许包含 α_i 或者 y_i 。

提示：仔细考虑一下那些由支持矢量 x_1 、 x_2 的决策规则产生的值。

w

$$x_1 - x_2 = [1 \ -1]$$

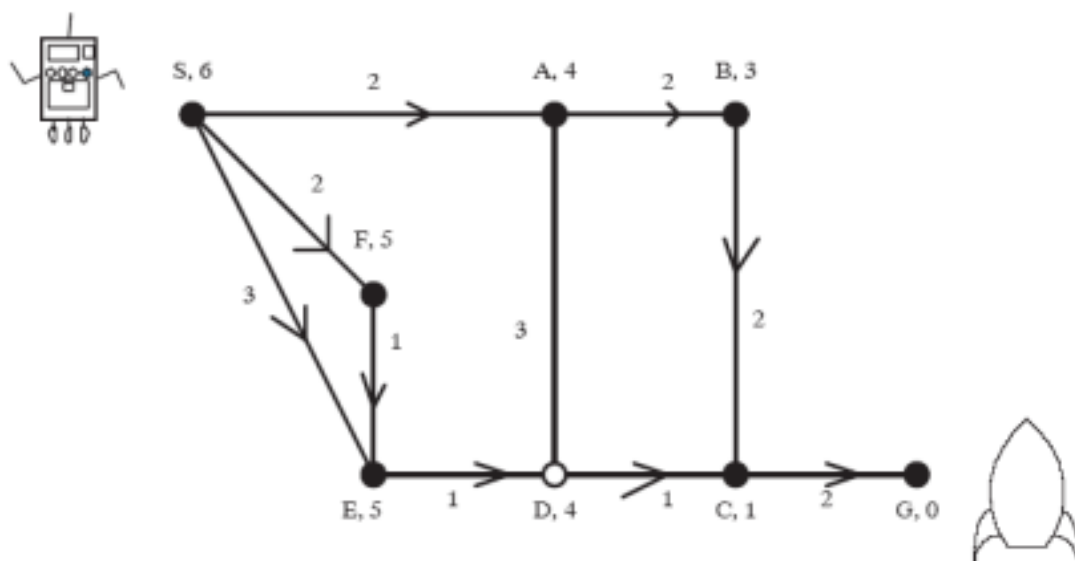
b

$$0$$

判决边界穿过了原点，所以 $b=0$

然后， $w \cdot x_1 = (a_1 x_1 + a_2 x_2) \cdot x_1 = a_1 x_1 \cdot x_1 = a_1$ 因为 a_1 和 a_2 是正交的，并且 x_1 是一个单位向量。另外，因为 x_1 是正的支持矢量，所以 $w \cdot x_1 = 1$ 。因此 $a_1 = 1$ 。同理有 $a_2 = -1$ ；

草稿，可撕下：



草稿，可撕下：

约束条件：

约束条件 1	一场舞不会在多于一个时间段中出现
约束条件 2	在时间段 1 和 6 中的舞蹈要特别精彩
约束条件 3	如果同一个演员参加了两个或两个以上的舞蹈，那么这些舞蹈不能连在一起

数据：

下面这个表格是关于哪些节目共享了演员的说明：

	A	B	C	D	E	F
A			Shares			
B			Shares		Shares	
C	Shares	Shares		Shares		
D			Shares			Shares
E		Shares				Shares
F				Shares	Shares	

类型转换的扣分表：

	芭蕾	hip hop	爵士	摇摆舞
芭蕾	10	5	1	2
Hip Hop	5	10	1	2
爵士	1	1	10	3
摇摆舞	2	2	3	10

舞蹈的类型

Dance	A	B	C	D	E	F
Genre	Hip Hop	Hip Hop	爵士	摇摆舞	芭蕾舞	爵士

节目精彩表程度表：

舞蹈	是否很精彩
A	Yes
B	
C	Yes
D	Yes
E	
F	

草稿，可撕下：

负 sigmoid 值：

z	$n(z)$	z	$n(z)$
-3.00	-0.05	0.25	-0.56
-2.75	-0.06	0.50	-0.62
-2.50	-0.08	0.75	-0.68
-2.25	-0.10	1.00	-0.73
-2.00	-0.12	1.25	-0.78
-1.75	-0.15	1.50	-0.82
-1.50	-0.18	1.75	-0.85
-1.25	-0.22	2.00	-0.88
-1.00	-0.27	2.25	-0.90
-0.75	-0.32	2.50	-0.92
-0.50	-0.38	2.75	-0.94
-0.25	-0.44	3.00	-0.95
0.00	-0.50	3.25	-0.96

实现神经网络梯度下降的有效方法：

下面的公式有两个前提：

规则的 sigmoid 单元： $s(z) = \frac{1}{1 + e^{-z}}$

误差计算公式： $E = \frac{1}{2} \sum (y^* - y)^2$

下降规则

$$w_{i \rightarrow j} = w_{i \rightarrow j} - r \frac{dE}{dw_{i \rightarrow j}} = w_{i \rightarrow j} - r \delta_j y_i$$

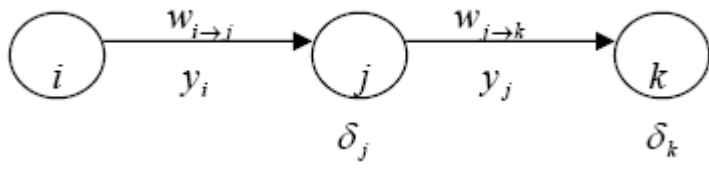
$$\delta_j = \frac{ds(z_j)}{dz_j} \sum_k \delta_k w_{j \rightarrow k}$$

反向传播规则：

- 1、将权值初始化为比较小的随机值；
- 2、选择一个随机样品的输入特征矢量；
- 3、为每个单元（前向传播）计算所有的输入值（ z_j ）和输出值（ y_j ）

$$\delta_n = \frac{ds(z_n)}{dz_n} (y_n - y_n^*) = y_n (1 - y_n) (y_n - y_n^*)$$

- 4、为输出层计算 δ_n 值，
- 5、按反向传播规则的前一层计算 δ_i 值（对每一层都重复计算）
- 6、按下降规则计算权值的变化量（对每个权值都重复计算）



y_i 是 x_i 的输入层