试卷一答案：

一、解：

C　B　AD　D　A

二、解：

由 $e_{ij}=\frac{count(\mathrm{A}=\mathrm{a}_i)\times count(B=b_i)}{n}$ 得：

$$e_{11}=\frac{300\times 450}{1500}=90 \qquad e_{12}=\frac{1200\times 450}{1500}=360$$

$$e_{21}=\frac{300\times 1050}{1500}=210 \qquad e_{22}=\frac{1200\times 1250}{1500}=840$$

所以

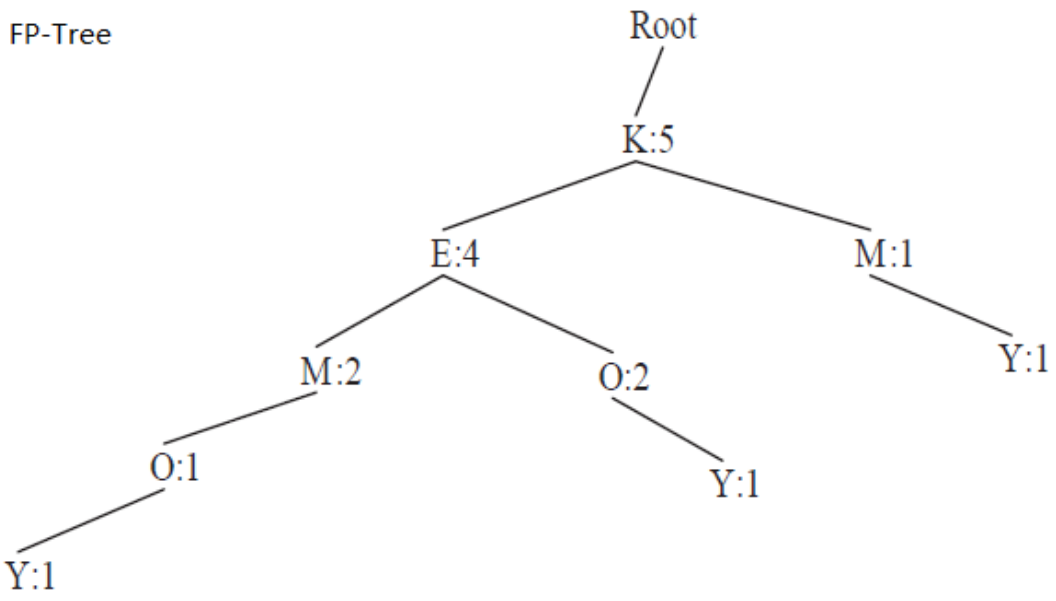
$$\chi^2=\frac{(250-90)^2}{90}+\frac{(50-210)^2}{210}+\frac{(200-360)^2}{360}+\frac{(1000-840)^2}{840}$$
$$=284.44+121.90+71.11+30.48=507.93$$

三、解：

分类和回归的区别在于输出变量的类型。定量输出称为回归，或者说是连续变量预测；定性输出称为分类，或者说是离散变量预测。

四、解：

（FP-Tree（不唯一））



item	conditional pattern base	conditional tree	frequent pattern
y	{ {k,e,m,o:1}, {k,e,o:1}, {k,m:1} }	k:3	{k,y:3}
o	{ {k,e,m:1}, {k,e:2}	k:3,e:3	{k,o:3}, {e,o:3}, {k,e,o:3}
m	{ {k,e:2}, {k:1} }	k:3	{k,m: 3}
e	{ {k:4} }	k:4	{ k,e:4 }

五、解：

（1）在数据集中有 20 个正样本和 500 个负样本，因此在根结点处错误率为

E=1-max（$\frac{20}{520}$，$\frac{500}{520}$）=$\frac{20}{520}$

如果按照属性 X 划分，则：

	X=0	X=1	X=2
+	0	10	10
-	200	0	300

$Ex=_0=0/310=0$　　$Ex=_1=0/10=0$　　$Ex=_2=10/310$

$$\triangle x = E - \frac{200}{520}*0 - \frac{10}{520}*0 - \frac{310}{520}*\frac{10}{310} = \frac{10}{520}$$

如果按照属性 Y 划分，则：

	Y=0	Y=1	Y=2
+	0	20	0
-	200	100	200

$E_{Y=0}=0/200=0$　　$E_{Y=1}=20/120$　　$E_{Y=2}=0/200=0$

$$\triangle x = E - \frac{120}{520}*\frac{20}{120} = 0$$

因此 X 被选为第一个分裂属性，因为 X=0 和 X=1 都是纯节点，所以使用 Y 属性去分割不纯节点 X=2。

Y=0 节点包含 100 个负样本，Y=1 节点包含 10 个正样本和 100 个负样本，Y=2 节点 100 个负样本，所以子节点被标记为”_”。整个结果为：

类标记= { +，X=1
　　-,其他

（2）

		预测类	
		+	-
实际类	+	10	10
	-	0	500

accuracy: $\frac{510}{520}$=0.9808， aprecision: $\frac{10}{10}$=1.0

recall: $\frac{10}{20}$=0.5， F-measure: $\frac{2*0.5*1.0}{1.0+0.5}$=0.6666

（3）由题可得代价矩阵为

		预测类	
		+	-
实际类	+	0	500/20=2.5
	-	1	0

决策树在（1）之后还有 3 个叶节点，X=2∧Y=0, X=2∧Y=1, X=2∧Y=2。

其中 X=2∧Y=1 是不纯节点，误分类该节点为”+”类的代价为：10*0+100*1=100，误分该节点为“-“类的代价为：10*25+100*0=250。所以这些节点被标记为”+”类。

分类结果为：

$$\text{类标记}=\begin{cases}+ & X=1 \vee (X=2 \wedge Y=1)\\ - & \text{其他}\end{cases}$$

六、解：

比如第一次聚类，两个聚簇的中心坐标如下：

聚类	**中心坐标**	
	$\bar{X}_1$	$\bar{X}_2$
（A、B）		
（C、D）		

第二步：计算某个样品到各类中心的欧氏平方距离，然后将该样品分配给最近的一类。对于样品有变动的类，重新计算它们的中心坐标，为下一步聚类做准备。先计算 A 到两个类的平方距离：

$$d^2(A,(AB)) = (5-2)^2 + (3-2)^2 = 10$$

$$d^2(A,(CD)) = (5+1)^2 + (3+2)^2 = 61$$

由于 A 到（A、B）的距离小于到（C、D）的距离，因此 A 不用重新分配。计算 B 到两类的平方距离：

$$d^2(B,(AB)) = (-1-2)^2 + (1-2)^2 = 10$$
$$d^2(B,(CD)) = (-1+1)^2 + (1+2)^2 = 9$$

由于 B 到（A、B）的距离大于到（C、D）的距离，因此 B 要分配给（C、D）类，得到新的聚类是（A）和（B、C、D）。更新中心坐标如下表所示。

聚类	中心坐标	
	$\bar{X}_1$	$\bar{X}_2$
（A）	5	3
（B、C、D）	-1	-1

第三步：再次检查每个样品，以决定是否需要重新分类。计算各样品到各中心的距离平方，结果见下表。

聚类	样品到中心的距离平方			
	A	B	C	D
（A）	0	40	41	89
（B、C、D）	52	4	5	5

到现在为止，每个样品都已经分配给距离中心最近的类，因此聚类过程到此结束。最终得到 K=2 的聚类结果是 A 独自成一类，B、C、D 聚成一类。

七、解：

假警告率 =（99%*1%）/(99%*1%+1%*99%)=50%

检测率 =（1%99%）/(1%)=99%

八、解：

（1）图 1 中，对象 p，q，p1 是核心对象；

图 2 中，对象 p，q，o 是核心对象。因为他们的 -领域内斗至少包含 3(MinPts)个对象。

（2）图 1 中，对象 p 是从对象 p1 直接密度可达的，反之亦然；对象 q 是从对象 p1 直接密度可达的，反之亦然。

图 2 中，对象 p，q，o 中不存在直接密度可达的对象。

（3）图 1 中，对象 p，q，p1 相互间是密度可达的；

图 2 中，对象 p，q，o 相互间是密度可达的。

（4）图 1 中，对象 p，q，p1 相互间是密度相连的；

图 2 中，对象 p，q，o 相互间是密度相连的。

九、解：

（1）均值 2.29，方差 1.51

（2）离群点 24.0

试题二答案

一、解：

（1）$d = \sqrt{(20-22)^2 + (0-1)^2 + (36-42)^2 + (8-10)^2} = 3\sqrt{5}$ ……

（2）Hi~ o(*￣▽￣*)ブ ……

（3）$d = \sqrt[3]{(20-22)^3 + (0-1)^3 + (36-42)^3 + (8-10)^3} = \sqrt[3]{233}$ ……

（4）$d(i,j) = \max_{f}^{p} | x_{if} - x_{jf} |$，所以 d=max{2,1,6,2}=6 ……

二、解：

a) Hadoop

基于分布式文件系统 HDFS 的分布式批处理计算框架。适用于数据量大，SPMD(单程序多数据)的应用。

b) Spark

基于内存计算的并行计算框架。适用于需要迭代多轮计算的应用。

c) MPI

基于消息传递的并行计算框架。适用各种复杂应用的并行计算。支持 MPMD(多程序多数据) ，开发复杂度高

三、解：

最小支持度计数为 5 60%=3

Apriori:

C1 =

m	3
o	3
n	2
k	5
e	4
y	3
d	1
a	1
u	1
c	2
i	1

L1 =

m	3
o	3
k	5
e	4
y	3

C2 =

mo	1
mk	3
me	2
my	2
ok	3
oe	3
oy	2
ke	4
ky	3
ey	2

L2 =

mk	3
ok	3
oe	3
ke	4
ky	3

C3=

oke	3

L3=

oke	3

四、解：

< {1} {2} {3} {4} >
< {1} {2 5} {3} >
< {1} {5} {3 4} >
< {2} {3} {4} {5} >
< {2 5} {3 4} >

五、解：

（1）$Info(D) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.940$ ……

$$Info_{income}(D) = \frac{4}{14}\times(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}) + \frac{6}{14}\times(-\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6}) + \frac{4}{14}\times(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}) = 0.9111$$

$Gain(income) = Info(D) - Info_{income}(D) = 0.940 - 0.9111 = 0.029$ ……

（2）……

$$SplitInfo_{income}(D) = -\frac{4}{14}\times\log_2\frac{4}{14} - \frac{6}{14}\times\log_2\frac{6}{14} - \frac{4}{14}\times\log_2\frac{4}{14} = 1.557$$

$GainRatio(income) = Gain(income) / SplitInfo_{income}(D) = 0.029 / 1.557 = 0.019$

（3）……

$$Gini_{income\in\{low,medium\}}(D) = \frac{10}{14}Gini(D_1) + \frac{4}{14}Gini(D_2)$$
$$= \frac{10}{14}\left(1-\left(\frac{7}{10}\right)^2-\left(\frac{3}{10}\right)^2\right)+\frac{4}{14}\left(1-\left(\frac{2}{4}\right)^2-\left(\frac{2}{4}\right)^2\right)$$
$$= 0.443$$

六、解：

每个类的先验概率 $P(C_i)$为： ……

P(buys-computer=yes)=9/14

P(buys-computer=no)=5/14

条件概率为： ……

P(age=youth|buys-computer=yes)=2/9

P(age=youth|buys-computer=no)=3/5

P(income=medium|buys-computer=yes)=4/9

P(income=medium|buys-computer=no)=2/5

P(student=yes|buys-computer=yes)=6/9

P(student=yes|buys-computer=no)=1/5

P(credit=fair|buys-computer=yes)=6/9

P(credit=fair|buys-computer=no)=2/5

使用上面的概率，得到：

P(X| buys-computer=yes)= P(age=youth|buys-computer=yes)

×P(income=medium|buys-computer=yes)

×P(student=yes|buys-computer=yes)

×P(credit=fair|buys-computer=yes)=32/729 ……

类似的，

P(X| buys-computer=no)=12/625 ……

为了找出最大化 $P(X|C_i)P(C_i)$，计算

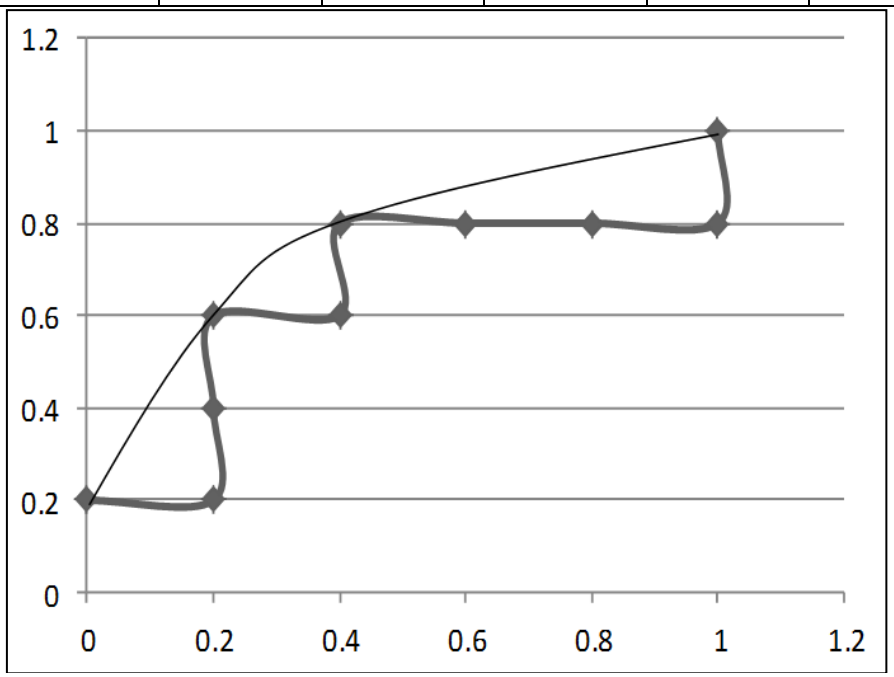
P(X| buys-computer=yes) P(buys-computer=yes)=16/567=0.028 ……

P(X| buys-computer=no) P(buys-computer=no)=6/875=0.007 ……

因此，对于元组 X，朴素贝叶斯分类预测元组 X 的类为 buys-computer=yes ……

七、解：

	TP	FP	TN	FN	TPR	FPR
P	1	0	5	4	0.2	0
N	1	1	4	4	0.2	0.2
P	2	1	4	3	0.4	0.2
P	3	1	4	2	0.6	0.2
N	3	2	3	2	0.6	0.4
P	4	2	3	1	0.8	0.4
N	4	3	2	1	0.8	0.6
N	4	4	1	1	0.8	0.8
N	4	5	0	1	0.8	1
P	5	5	0	0	1	1



八、解：k-均值和 k-中心点算法都可以进行有效的聚类。

（1）k-均值

优点： 高效，k-均值算法复杂度为 O(tkn)，n 是对象数目，k 是聚类数目，t 是迭代次数，一般的 k, t << n;

缺点： 1）局部最优解；

2）只适用于连续的固定的 n 维数据

3）需要先确定聚类数目 k；

4）对噪音和离群点比较敏感；

5）只适用于凸型数据聚类。

k-中心点

优点： 1）可适用于范围可变的数据；

2）能够处理对噪声或离群点。

缺点： 1）局部最优解

2）只适用于数据集较小的数据集，对较大的数据集不适用（计算的复杂性）算法复杂度为 O(k(n-k)2)。

3）需要先确定聚类数目 k；

4）只适用于凸型数据聚类

（2）层次化聚类方法

优点：没有局部极小问题或是很难选择初始点的问题

缺点：计算存储的代价昂贵。

试题三答案：

一、解：

B C A C C　　B C A A A　　B A B B D

二、解：

ABC　　ACD　　ABCD　　AD　　ABCD

三、解：

g(D,A) = H(4/10, 6/10) - 7/10 * H(4/7, 3/7) - 3/10 * H(1, 0)

g(D,B) = H(4/10, 6/10) - 4/10 * H(3/4, 1/4) - 6/10 * H(5/6, 1/6)

信息增益表示特征 X 使得类 y 的不确定性减少的程度

四、解：

这属于聚类分析。因为这是无监督的学习，事先不知道各个类别的标准。而分类分析属于有监督的学习，事先知道分类的标准。

两者的区别主要是数据中是否有类标号。 从数据方面来说，分类挖掘的数据必须有类标号，也就是有专家参与。

五、解：

（1）：平均值是 25.08 ， 20% 的截断均值（两端各去掉两个数以后的平均值）是 18.8 ，中位数是 19 。

（2）规范化后，转换后的值为（ 0 ， 1 ， 0.31 ）

（3）深度为 4 进行划分，得到三个箱 （ 6 ， 7 ， 9 ， 11 ），（ 12 ， 18 ， 20 ， 21 ），（ 25 ， 35 ， 37 ， 100 ）

边界值平滑后的结果为 （ 6 ， 6 ， 11 ， 11 ），（ 12 ， 12 ， 21 ， 21 ），（ 25 ， 25 ， 25 ， 100 ）

六、解：

欠拟合的原因：模型复杂度过低，不能很好的拟合所有的数据，训练误差大；

避免欠拟合：增加模型复杂度，如采用高阶模型（预测）或者引入更多特征（分类）等。

过拟合的原因：模型复杂度过高，训练数据过少，训练误差小，测试误差大；

避免过拟合：降低模型复杂度，如加上正则惩罚项，如 L1，L2，增加训练数据等。

七、解：

剑桥分析有三大法宝：心理学模型、大数据分析、定向广告。

首先，他们从各种地方获取个人数据，比如土地登记信息、汽车数据、购物数据、优惠券、俱乐部会员，以及 FB 账户信息等。再把这些信息与选民名册等大数据整合到一起，一起放进已研发出的心理学模型中，原先的数字足迹变成了完整又具体的大活人，他们有担忧、有需求、有兴趣、有癖好，还附带手机号码、信用卡类型、电子邮箱和家庭住址。接下来，你就可以根据自己的需求，向这些早已被你研究透的人们灌输思想了。

在川普和希拉里展开第三场电视辩论的时候，剑桥分析用川普的观点在 FB 上精心测试了

17.5 万个版本的广告，然后跟踪人们在网络上的举动和兴趣，恰如其分地投放 4 到 5 万条，不同版本的差别都仅仅是细节：比如标题、颜色、照片、视频……然后就是等待猎物的反馈：比如某人是宾夕法尼亚州一个摇摆不定的选民，他有没有点击关于希拉里邮件门的帖子呢？点了，就显示更多的内容，看看希拉里是如何失职的。没点？自动脚本就换个标题，或者换个切入点——比如这个人容易听信权威，标题就自动更正为：《情报部门高官一致认为：希拉里邮件门事件危及国家安全》。

总之就是反复向你投放他们想让你看到的内容，直到你最终被他们洗脑。

试题四答案

一、解：

C C D C C　A D A C D　C D C D A

二、解：

BD　ABCD　BCE　AD　ABC　D　ABD　C　ACD　A　BC　BC　BD　C　C

三、解：

（1）被评为垃圾邮件的发信邮件地址，信的内容，信的格式（长度，段落等）

（2）准确率为 1%，召回率为 1.01%

（3）对正样本进行上采样，或者对正样本加大的权重

四、解：

d （ p ， q ） =1+1+ （ 20-18 ） =4

d （ p ， C1 ） = （ 1-25/30 ） + （ 1-20/30 ） + （ 20-18 ） =2.5

d （ p ， C2 ） = （ 1-3/15 ） + （ 1-0/15 ） + （ 24-18 ） =7.8

d （ q ， C1 ） = （ 1-5/30 ） + （ 1-4/30 ） + （ 20-20 ） =1.7

d （ q ， C2 ） = （ 1-12/15 ） + （ 1-2/15 ） + （ 24-20 ） =5.07

d （ C1 ， C1 ） = （ 1- （ 25*3+5*12 ） /30*15 ）

+ （ 1- （ 6*1+4*2 ） /15*30 ） + （ 24-20 ） =5.67

五、解：

K-means 算法通过最小化平方距离，通过迭代发现 K 个聚簇，在每次迭代中，需要计算均值点，通过每个点与均值点的距离来重新调整聚类。

选择聚类中心通过计算每个聚簇中所有点在每个维度的平均值来获得。

试题五答案

一、解：

1.离群点可以是合法的数据对象或者值。 （T）

2.离散属性总是具有有限个值。 （F）

3.关联规则挖掘过程是发现满足最小支持度的所有项集代表的规则。 （F）

4.K 均值是一种产生划分聚类的基于密度的聚类算法，簇的个数由算法自动地确定。（F）

5.如果一个对象不属于任何簇，那么该对象是基于聚类的离群点。 （T）

二、解：

P(青年 | 购买) = 2/9 = 0.222

P(收入中等 | 购买) = 4/9 = 0.444

P(学生 | 购买) = 6/9 = 0.667

P(信用中 | 购买) = 6/9 = 0.667

P(X | 购买) = 0.222 ×0.444 ×0.667 ×0.667=0.044

P(青年 | 不买) = 3/5 = 0.6

P(收入中等 | 不买) = 2/5 = 0.4

P(学生 | 不买) = 1/5 = 0.2

P(信用中 | 不买) = 2/5 = 0.4

P(X | 不买) = 0.6 ×0.4 ×0.2 ×0.4=0.019

$$P(C_{买})=9/14=0.643$$

$$P(C_{不买})=5/14=0.357$$

P(购买|X) = 0.044 ×0.643

=0.028（1 分）

P(不买|X) = 0.019 ×0.357

=0.007（1 分）

试题六答案

第一题. 单选题

AACBA　　CDBAD　　CCDCB　　ACBCC

第二题. 多选题

AB　　AD　　ABCDE　　ABCDE　　BD

第三题. 判断题

TTTFT　　FTFTF　　FFTFT　　FFTFF

试题七答案

第一题. 单选题

ABADB　　CABAA　　ABDCD　　CADDD　　CDACB　　DACCC

第二题. 多选题

CD　　BC　　ABCD　　AD　　AB　　AC　　ACD　　BCDE　　ABCD　　BCD

试题八答案

第一题. 单选题

ABACB BDCBB AAAAB ACBCB

第二题. 多选题

ABC ABCD ABC AB ABD BC BC ABCD ABC AB

BCD ABC AB ABC BCD

试题九答案

一、单选题

BACDC BCADA ABCDC ABBBC BBACC CCDAC

二、不定项选择题

CD ABC AC BCD ABC ACD ABD D AC ACD

三、

Support(看乒乓球→看篮球) = 2000 / 5000 = 40%

Confidence(看乒乓球→看篮球) = 2000 / 3000 = 66.7%

$$lift = \frac{P(\text{看乒乓球} \cup \text{看篮球})}{P(\text{看乒乓球})P(\text{看篮球})} = \frac{2000/5000}{3000/5000 * 3750/5000} = 0.89$$

四、

$$P(h+|D) = P(h+) * \frac{P(D|h+)}{P(D)}$$

$$= 0.2 * \frac{(0.3*0.2*0.2)}{P(D)} = 0.0096/P(D)$$

$$P(h-|D) = P(h-) * \frac{P(D|h-)}{P(D)}$$

$$= 0.8 * \frac{(0.01*0.01*0.2)}{P(D)} \text{（2 分）} = 0.000016/P(D)$$

$P(h+|D) > \mathrm{P(h-|D)}$

答：该邮件是垃圾邮件

五、

$\mathrm{c(ABC \rightarrow D) \geq c(AB \rightarrow CD) \geq c(A \rightarrow BCD)}$

c(ABC→D)= support({A}∪{B}∪{C}∪{D})/ support({A}∪{B}∪{*C*})

c(AB → CD)= support({A}∪{B}∪{C}∪{D})/support({A}∪{*B*})

c(A → BCD) = support({A}∪{B}∪{C}∪{D})/ support({*A*})

很显然：

support({A}∪{B}∪{*C*})≤ support({A}∪{*B*})≤ support({*A*})

因此：c(ABC → D) ≥ c(AB → CD) ≥ c(A → BCD)

六、

神经网络使用激活函数能够增加模型的非线性映射，提高网络的拟合和表达能力；

$$; f'(x) = \{(1+e^{-x})^{-1}\}'$$

$$= -(1+e^{-x})^{-2}(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= f(x)\frac{e^{-x}}{1+e^{-x}}$$

$$= f(x)\left(1-\frac{1}{1+e^{-x}}\right)$$

$$= f(x)(1-f(x))$$

$$f'(0) = f(0)(1-f(0)) = \frac{1}{4}$$

七：

问题 1 解答：

小梅采用 OneHotEncoder 独热编码，而小雷采用的是 LabelEncoder 标签编码（即用某一个数字代表一种类型，如 1 代表一线城市，2 代表二线城市，3 代表三线城市）。如果模型损失函数对变量的数值大小是敏感的，如 SVM、LR、GLM 等，为模型 A；如果模型损失函数对变量数据大小不敏感，数值的存在意义是为了排序，如各种树模型，则为模型 B。显然该题用的 LR 模型对变量数值大小是敏感的，所以小梅的编码方式更合适。

问题 2 解答：

beta 为机器学习模型中的截距，如果设置为 1，与事实相比过大，可能需要模型训练更长时间。所以小梅更好，能在短时间找到最优的模型参数。

问题 3 解答：

在训练样本中拟合的很好，但是在测试样本中效果比较差，属于过拟合问题。该损失函数使用的是经验风险最小化，不是结构风险最小化，泛化能力差，容易过拟合。（结构风险=经验风险+置信风险，置信风险是一个减函数，整个公示反映了经验风险和真实误差的差距上界，表征了根据经验风险最小化原则得到的模型的泛化能力。称为泛化误差上界。）

问题 4 解答：

AUC 最大的应用应该就是点击率预估（CTR）的离线评估。其计算过程如下：

得到结果数据，数据结构为：（输出概率，标签真值）；

对结果数据按输出概率进行分组，得到（输出概率，该输出概率下真实正样本数，该输出概率下真实负样本数）。这样做的好处是方便后面的分组统计、阈值划分统计等；

对结果数据按输出概率进行从大到小排序；

从大到小，把每一个输出概率作为分类阈值，统计该分类阈值下的 TPR 和 FPR；

微元法计算 ROC 曲线面积、绘制 ROC 曲线。

试题十答案

一、单选题

ADDBD　CABDC　CBBCB　CCAAC　DDCCC　CCAAB

二、判断题

FFTTF　FTFTF

三、不定项选择题

BCD　ABC　ABD　ABC　ABC

四、

答：聚类算法主要有：层次的方法（hierarchical method）、划分方法（partitioning method）、基于密度的方法（density-based method）、基于网格的方法(grid-based method)、基于模型的方法（model-based method）等。其中，前两种算法是利用统计学定义的距离进行度量。

K-Means 算法的计算原理如下：首先从 n 个数据对象任意选择 k 个对象作为初始聚类中心；

而对于所剩下其它对象，则根据它们与这些聚类中心的相似度（距离），分别将它们分配给与其最相似的（聚类中心所代表的）聚类；然后再计算每个所获新聚类的聚类中心（该聚类中所有对象的均值）；不断重复这一过程直到标准测度函数开始收敛为止。

在算法中，一般都采用均方差作为标准测度函数，算法收敛后各聚类本身尽可能的紧凑，而各聚类之间尽可能的分开。

五、

答：在模型建立的时候，如果一味的追求提高对训练数据的预测能力，所选模型的复杂度往往会比真实的模型高，这种现象称之为过拟合。从原理上来说，过拟合是对训练数据的过度学习，得到的模型参数太多太复杂，所建立模型太过于依赖训练数据，从而导致模型放在预测数据上时反而得不到很好的效果。因此在模型建立和选择时，不仅仅要考虑在训练集上准确率高，更重要的是在测试集上的准确性。

防止过拟合最常用的方法就是模型的正则化，即在模型的经验风险后面加上一个正则项（惩罚项），正则项一般是模型复杂度的单调递增函数，模型越复杂，正则项也越大。通过添加正则项强迫机器去学习尽可能简单的模型。正则化的作用就是选择经验风险和模型复杂度都比较小的模型。正则化符合奥卡姆剃刀原则：在所有可以选择的模型中，能够很好地解释已知数据同时十分简单的模型才是最好的模型。

六、

1）年龄均值=

(23+23+27+27+39+41+47+49+50+52+54+54+56+57+58+58+60+61)/18=836/18=46.44

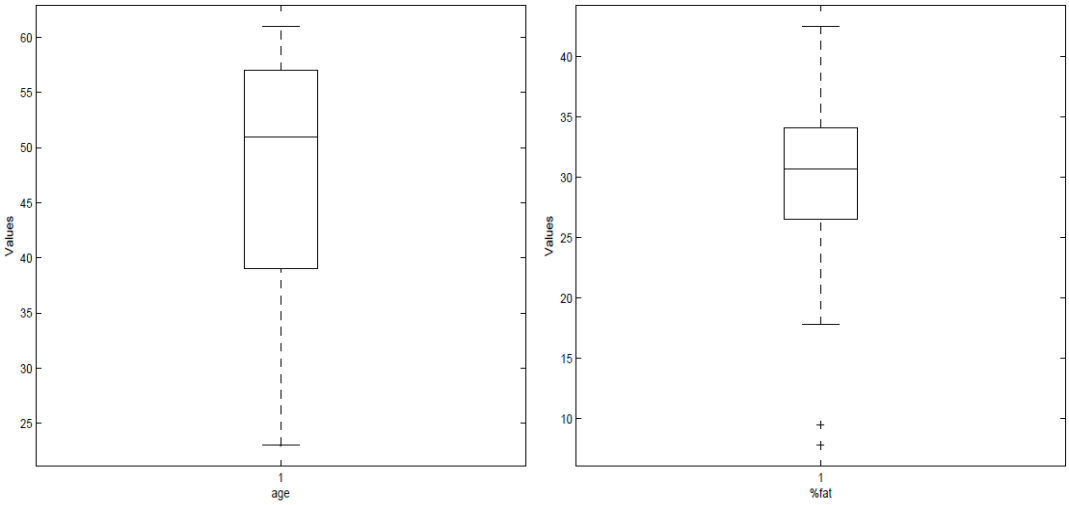
年龄中位数=(50+52)/2=51

标准差=方差的平方根=开根号（$1/n-1[\sum(Xi)^2-1/n-1(\sum Xi)^2]$）注意这里是抽样（即估算样本方差），根号内除以（n-1）

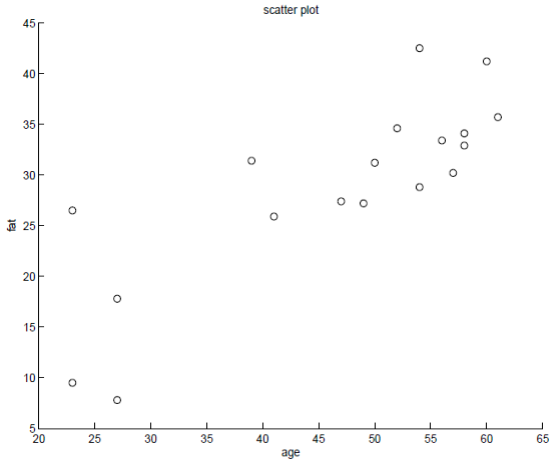
=开根号 1/17[2970.44]=13.22

脂肪百分比均值=28.78, 中位数=30.7, 标准差= 9.25

2）绘制年龄和脂肪百分比的盒图



3）根据这两个属性，绘制散布图



4）根据 z-score 规范化来规范化这两个属性

age	23	23	27	27	39	41	47	49	50
z-age	-1.83	-1.83	-1.51	-1.51	-0.58	-0.42	0.04	0.20	0.28
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
z-%fat	-2.14	-0.25	-2.33	-1.22	0.29	-0.32	-0.15	-0.18	0.27
age	52	54	54	56	57	58	58	60	61
z-age	0.43	0.59	0.59	0.74	0.82	0.90	0.90	1.06	1.13
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7
z-%fat	0.65	1.53	0.0	0.51	0.16	0.59	0.46	1.38	0.77

5）计算得到相关系数为 0.82 公式如下，两个属性变量呈正相关

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

$$r = \frac{1}{(n-1)} \sum \frac{(X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y}$$

七：

1）节点 4、5、6 的净输入 I_j 和输出 O_j 为：

单元j	净输入I_j	输出O_j
4	0.2+0 – 0.5–0.4 = –0.7	$1+(1+e^{0.7}) = 0.332$
5	– 0.3+0+0.2+0.2 = 0.1	$1+(1+e^{-0.1}) = 0.525$
6	(-0.3)(0.332)+(-0.2)(0.525)+0.1 = -0.105	$1+(1+e^{-0.105}) = 0.474$

2）节点 4、5、6 的误差 Err_j 为：

单元j	Err_j
6	(0.474)(1-0.474)(1-0.474) = 0.1311
5	(0.525)(1-0.525)(0.1311)(-0.2) = -0.0065
4	(0.332)(1-0.332)(0.1311)(-0.3) = -0.0087

3）更新后的权值和偏置为：

权或偏置	新值
w_{46}	-0.3 + (0.9)(0.1311)(0.332) = -0.261
w_{56}	-0.2 + (0.9)(0.1311)(0.525) = -0.138
w_{14}	0.2 + (0.9)(-0.0087)(1) = 0.192
w_{15}	-0.3 + (0.9)(0.0065)(1) = -0.306
w_{24}	0.4 + (0.9)(-0.0087)(0) = 0.4
w_{25}	0.1+ (0.9)(-0.0065)(0) = 0.1
w_{34}	-0.5 + (0.9)(-0.0087)(1) = -0.508
w_{35}	0.1+ (0.9)(-0.0065)(1) = 0.194
θ_6	0.1+(0.9)(0.1311) = 0.218
θ_5	0.2+(0.9)(-0.0065) = 0.194
θ_4	-0.4+(0.9)(-0.0087) = -0.408

4）根据链式法则，如果每一层神经元对上一层的输出的偏导乘上权重结果都小于 1 的话，那么即使这个结果是 0.99，在经过足够多层传播之后，误差对输入层的偏导会趋于 0，简言之，随着网络层数的增加，误差反向传播的梯度更新信息会朝着指数衰减的方式减少，这就是梯度消失。