

共六大题，基本都在老师考前画的重点内，不过还是有些出乎意料的…可能因为往年开卷的原因，感觉复习资料特别少，简单整理一下供大家参考。

## 一、概念题

- 1、机器学习过程，每个环节主要操作
- 2、集成学习的概念，BOOSTING 和 BAGGING 思想
- 3、交叉验证法
- 4、特征提取和特征选择的区别
- 5、隐马尔科夫模型定义形式，以及解决的三个问题？
- 6、贝叶斯决策过程

## 二、svm

- 1、svm 概念,其目的,什么是最优化分类面
- 2、最优拉格朗日形式
- 3、核函数为什么要引入函数
- 4、验证核函数，基本如下，只是  $x, y$  换成了  $x_1, x_2$ ，映射函数是  $(x^2 \ x \ 1/2)^T$

我们现在考虑核函数  $K(v_1, v_2) = \langle v_1, v_2 \rangle^2$ ，即“内积平方”。

这里面  $v_1 = (x_1, y_1)$ ,  $v_2 = (x_2, y_2)$  是二维空间中的两个点。

这个核函数对应着一个二维空间到三维空间的映射，它的表达式是：

$$\Phi(v) = v = \begin{bmatrix} x \\ y \end{bmatrix} \quad \Phi(v) = \begin{bmatrix} x^2 \\ \sqrt{2}xy \\ y^2 \end{bmatrix}$$

可以验证，

$$\begin{aligned} \langle \Phi(v_1), \Phi(v_2) \rangle &= \langle (x_1^2, \sqrt{2}x_1y_1, y_1^2), (x_2^2, \sqrt{2}x_2y_2, y_2^2) \rangle \\ &= x_1^2x_2^2 + 2x_1x_2y_1y_2 + y_1^2y_2^2 \\ &= (x_1x_2 + y_1y_2)^2 \\ &= \langle v_1, v_2 \rangle^2 \\ &= K(v_1, v_2) \end{aligned}$$

<http://blog.csdn.net/Weyoung>

## 三、神经网络、

1. 描述 bp 算法
2. 前向传播表达式 a1
3. 输出层输出形式  
(嗯，又一个原封不动的)

## 前向传播

$$a_1^{(2)} = f(W_{11}^{(1)}x_1 + W_{12}^{(1)}x_2 + W_{13}^{(1)}x_3 + b_1^{(1)})$$

$$a_2^{(2)} = f(W_{21}^{(1)}x_1 + W_{22}^{(1)}x_2 + W_{23}^{(1)}x_3 + b_2^{(1)})$$

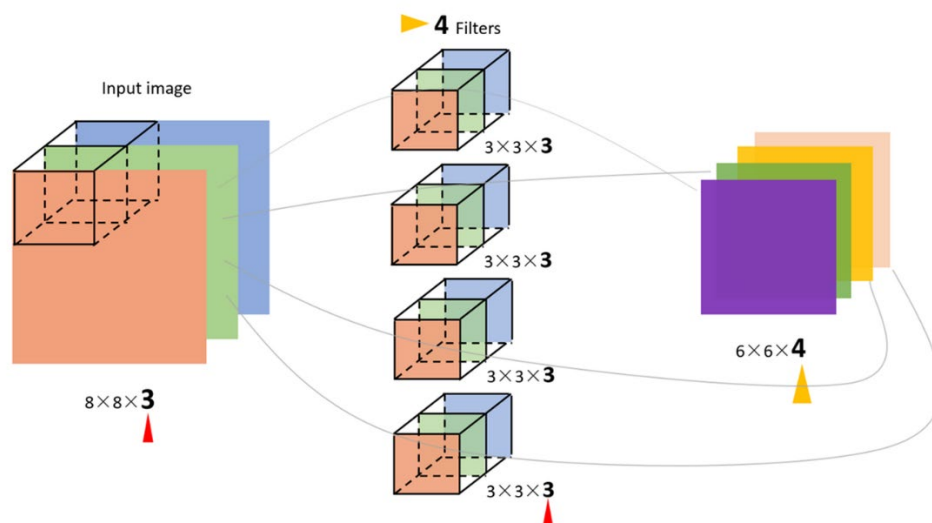
$$a_3^{(2)} = f(W_{31}^{(1)}x_1 + W_{32}^{(1)}x_2 + W_{33}^{(1)}x_3 + b_3^{(1)})$$

$$h_{W,b}(x) = a_1^{(3)} = f(W_{11}^{(2)}a_1^{(2)} + W_{12}^{(2)}a_2^{(2)} + W_{13}^{(2)}a_3^{(2)} + b_1^{(2)})$$

$$f \left( \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} & b_1^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} & b_2^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} & b_3^{(1)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} a_1^{(2)} \\ a_2^{(2)} \\ a_3^{(2)} \end{bmatrix}$$

#### 四、深度学习

1. 卷积层作用
2. 激活函数以及形式



#### 3. 输出

给出上图左 1，左二，问你卷积后输出几个，大小为多少，（输出即右一图，这页 ppt 我根本就沒瞅啊，现在就是后悔!!!）

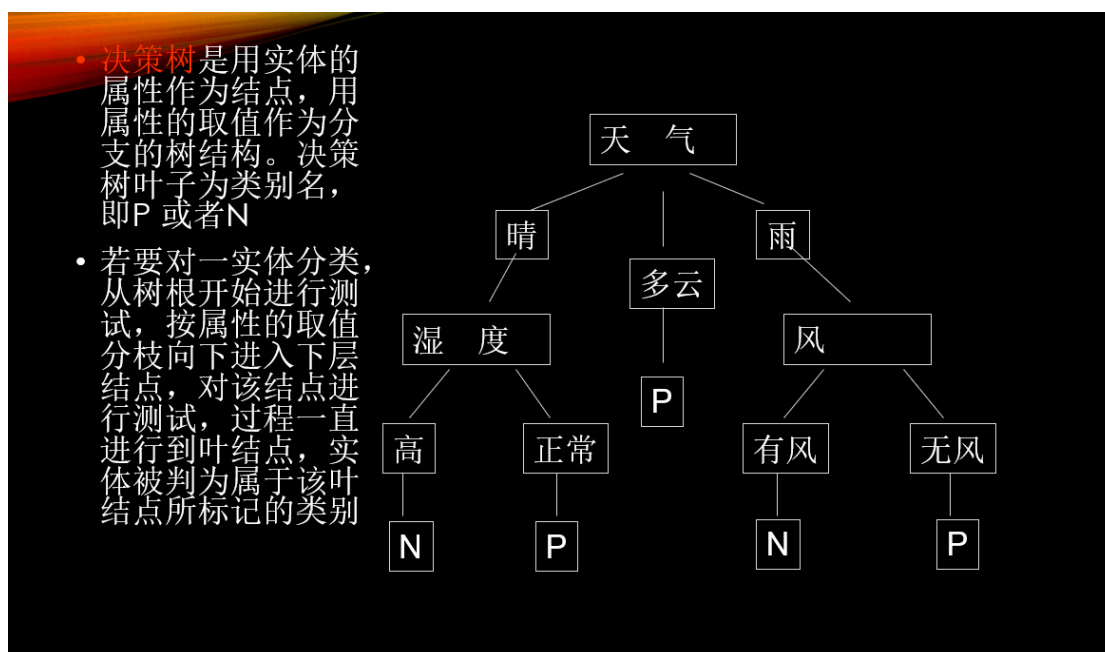
#### 五、决策树

1. 决策树算法思想是什么，两个分类
2. 给出一个表格
  - 1) 写出预处理数据集
  - 2) 决策树缺失项 以及依据
  - 3) 决策树分类规则
  - 4) 给数据判断结果

给出的表格、决策树基本如下：

NO.	属性				类别
	天气	气温	湿度	风	
1	晴	热	高	无风	N
2	晴	热	高	有风	N
3	多云	热	高	无风	P
4	雨	适中	高	无风	P
5	雨	冷	正常	无风	P
6	雨	冷	正常	有风	N
7	多云	冷	正常	有风	P
8	晴	适中	高	无风	N
9	晴	冷	正常	无风	P
10	雨	适中	正常	无风	P
11	晴	适中	正常	有风	P
12	多云	适中	高	有风	P
13	多云	热	正常	无风	P
14	雨	适中	高	有风	N

实体  
(样本)、  
概念的正  
例和反例、  
训练集



## 六、DBSCAN 算法应用计算（给距离矩阵求聚类结果）

下面给出一个样本事务数据库（见左表），对它实施DBSCAN算法。  
根据所给的数据通过对其进行DBSCAN算法，以下为算法的步骤（设 $n=12$ ，  
用户输入  $\varepsilon=1$ ，MinPts=4）

样本事务数据库

序号	属性 1	属性 2
1	1	0
2	4	0
3	0	1
4	1	1
5	2	1
6	3	1
7	4	1
8	5	1
9	0	2
10	1	2
11	4	2
12	1	3

距离矩阵

	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	3	0										
3	1.4	4.1	0									
4	1	3.2	1	0								
5	1.4	2.2	2	1	0							
6	2.2	1.4	3	2	1	0						
7	3.2	1	4	3	2	1	0					
8	4.1	1.4	5	4	3	2	1	0				
9	2.2	4.5	1	1.4	2.2	3.2	4.1	5.1	0			
10	2	3.6	1.4	1	1.4	2.2	3.2	4.1	1	0		
11	3.6	2	4.1	3.2	2.2	1.4	1	1.4	4	3	0	
12	3	4.2	2.2	2	2.2	2.8	3.6	4.5	1.4	1	3.2	0

聚出的类为{1, 3, 4, 5, 9, 10, 12}，{2, 6, 7, 8, 11}。

步骤	选择的点	在 $\varepsilon$ 中点的个数	通过计算可达点而找到的新簇
1	1	2	无
2	2	2	无
3	3	3	无
4	4	5	簇 $C_1$ : {1, 3, 4, 5, 9, 10, 12}
5	5	3	已在簇 $C_1$ 中
6	6	3	无
7	7	5	簇 $C_2$ : {2, 6, 7, 8, 11}
8	8	2	已在簇 $C_2$ 中
9	9	3	已在簇 $C_1$ 中
10	10	4	已在簇 $C_1$ 中
11	11	2	已在簇 $C_2$ 中
12	12	2	已在簇 $C_1$ 中

第1步，在数据库中选择一点1，由于在以它为圆心的，以1为半径的圆内包含2个点（小于4），因此它不是核心点，选择下一个点。

第2步，在数据库中选择一点2，由于在以它为圆心的，以1为半径的圆内包含2个点，因此它不是核心点，选择下一个点。

第3步，在数据库中选择一点3，由于在以它为圆心的，以1为半径的圆内包含3个点，因此它不是核心点，选择下一个点。

第4步，在数据库中选择一点4，由于在以它为圆心的，以1为半径的圆内包含5个点，因此它是核心点，寻找从它出发可达的点（直接可达4个：1, 3, 5, 10，间接可达3个），聚出的新类{1, 3, 4, 5, 9, 10, 12}，选择下一个点。

第5步，在数据库中选择一点5，已经在簇1中，选择下一个点。

第6步，在数据库中选择一点6，由于在以它为圆心的，以1为半径的圆内包含3个点，因此它不是核心点，选择下一个点。

第7步，在数据库中选择一点7，由于在以它为圆心的，以1为半径的圆内包含5个点，因此它是核心点，寻找从它出发可达的点，聚出的新类{2, 6, 7, 8, 11}，选择下一个点。

第8步，在数据库中选择一点8，已经在簇2中，选择下一个点。

第9步，在数据库中选择一点9，已经在簇1中，选择下一个点。

第10步，在数据库中选择一点10，已经在簇1中，选择下一个点。

第11步，在数据库中选择一点11，已经在簇2中，选择下一个点。

第12步，选择12点，已经在簇1中，由于这已经是最后一点所有点都已处理，程序终止。

最后，感谢我晨哥的试题回忆服务。整理不易，欢迎打赏（手动狗头）

[https://blog.csdn.net/weixin\\_42039835/article/details/103478413](https://blog.csdn.net/weixin_42039835/article/details/103478413)