一、网络空间安全智能决策-----周五14：00-16：00

实验一、word文档中红色字体Webshell

考点：

第一章：

1、机器学习的最终目的就是使得计算机拥有和人类一样的学习能力

2、网络安全一直和（人工智能）相伴相生...人们就一直识图使用（自动化）的方式去解决安全问题

3、WebShell检测

WebShell就是以**ASP**、**PHP**、**JSP**或者**CGI**等网页（**文件形式**）存在的一种命令执行环境，也可以将其成为一种网页后门。"WEB"的含义是需要服务器（**提供Web服务**），“Shell”的含义是取得对服务器的某种程度的（**操作权限**）。WebShell常常被入侵者利用，通过网站服务端口对网站服务器获取某种程度的操作权限。

4、网络安全案例：

恶意程序识别、骚扰短信识别、反信用卡欺诈、Linux后门检测、智能扫描器、DGA域名检测、验证码识别

5、如何让机器可以真正学会自动识别安全威胁

机器学习

6、机器学习分类

1）监督学习-->有标签样本

2）无监督学习-->无标签样本

7、计算学习理论是机器学习的理论基础，最重要的理论模型：

PAC：概率近似正确

$$P(|f(x)-y| \leq \epsilon) \geq 1-\delta$$

第二章

1、线性回归：擅长处理数值属性（连续）

若属性值之间存在“序”关系，可通过连续化将其转化为**连续值**

若属性值之间不存在“序”关系，则假定有K个属性值，通常转化为**K维向量。**

2、线性模型优化目标：均方误差最小化

令均方误差最小化，有 $(w^*, b^*) = \arg\min_{(w,b)} \sum_{i=1}^{m} (f(x_i) - y_i)^2$

$$= \arg\min_{(w,b)} \sum_{i=1}^{m} (y_i - wx_i - b)^2$$

对 $E_{(w,b)} = \sum_{i=1}^{m} (y_i - wx_i - b)^2$ 进行最小二乘参数估计

第三章

1、决策树核心思想：

是一种**监督学习算法**，广泛应用于分类和回归任务，通过从顶部的**根节点**到底部的**叶节点**，根据数据特征的递归分割来建模决策，每个**内部节点**代表一个属性，每个**叶节点**代表最终的**预测结果。**

2、决策树模型结构

决策树由根节点、内部节点、叶节点和分支组成

3、决策树发展史

第一个决策树算法：CLS

使决策树收到关注、成为机器学习主流技术的算法：ID3

最常用的决策树算法：C4.5

可用于回归任务的决策树算法：CART

基于决策树的最强大算法：RF

4、决策树算法的基本流程遵循“分而治之”的策略

5、信息熵表示随机变量不确定性的度量，即度量样本集合纯度。

6、ID3算法中----------重点

衡量哦我们选择某个属性进行划分时信息熵的变化，以属性a对数据D进行划分所得的信息增益。

7、剪枝是决策树对付“过拟合”的主要手段

8、决策树剪枝的基本策略

预剪枝：提前终止某些分支的生长

后剪枝：生成一颗完全树，再“回头剪枝

第四章

1、支持向量机的目标

决策边界离到它最近的那个点距离越大越好（即最大化决策边界的间隔）

2、支持向量机核函数解决线性不可分

将样本从原始空间映射到了一个高维空间，使得样本再这个高维空间内线性可分（即再高维空间中找到最优决策边界）然后再SVM求解。

3、常用核函数

线性核

多项式核

径向基函数

Sigmoid核

注意：绝对值核是不常用的！！！

第五章 神经网络

1、神经网络（模仿）的是人脑中的（神经元）结构，然后通过激活函数处理产生神经元的输出神经网络所学得的知识蕴含在连接（权重）与（阈值）中

2、激活函数的作用是引进非线性，帮助网络学习复杂的函数

3、多层网络：包含隐层的网络

由输入层、隐藏层和输出层组成

4、前馈网络

由输入层、隐藏层和输出层组成。神经元之间不存在（同层）连接也不存在（跨层）连接，即网络中无环路或者回路。

5、隐层和输出层神经元亦称"功能单位"。其中隐藏层的主要作用是增加网络的（非线性能力）。

6、误差逆传播算法

进行优化：逆传播/反向传播算法是最成功、最常用的神经网络学习算法。

7、Softmax分类器（计算）

Softmax 分类器 要计算

- 如何把一个得分值转换成一个概率值？
- 将得分值归一化

$$\mathrm{Softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

假设神经网络的输出层有三个神经元，输出：$z = [z_1, z_2, z_3] = [2, 1, 0.5]$

步骤 1: 计算每个分数的指数

$$e^{z_1} = e^2$$
$$e^{z_2} = e^1$$
$$e^{z_3} = e^{0.5}$$

步骤 2: 计算指数的总和

$$\sum_j e^{z_j} = e^2 + e^1 + e^{0.5}$$

步骤 3: 每个类别的softmax概率

$$\mathrm{Softmax}(z_1) \approx \frac{7.389}{11.756} \approx 0.628$$
$$\mathrm{Softmax}(z_2) \approx \frac{2.718}{11.756} \approx 0.231$$
$$\mathrm{Softmax}(z_3) \approx \frac{1.649}{11.756} \approx 0.140$$

- 类别 1 的概率：≈ 0.628
- 类别 2 的概率：≈ 0.231
- 类别 3 的概率：≈ 0.140

第六章 贝叶斯分类器

- 求解：穿长裤的人里面有多少女生
- 穿长裤的总人数：

U * P(Boy) * P(Pants|Boy) + U * P(Girl) * P(Pants|Girl)

- 穿长裤的是女生的概率：P(Girl|Pants)？

穿长裤-女生-人数 / 穿长裤-总人数

$$P(\mathrm{Girl}|\mathrm{Pants}) = \frac{U * P(\mathrm{Girl}) * P(\mathrm{Pants}|\mathrm{Girl})}{U * P(\mathrm{Boy}) * P(\mathrm{Pants}|\mathrm{Boy}) + U * P(\mathrm{Girl}) * P(\mathrm{Pants}|\mathrm{Girl})}$$

贝叶斯公式推导

➢ 与总人数无关

$$P(Girl|Pants) = \frac{P(Girl) * P(Pants|Girl)}{P(Boy) * P(Pants|Boy) + P(Girl) * P(Pants|Girl)}$$

$P(Boy) * P(Pants|Boy) + P(Girl) * P(Pants|Girl) = P(Pants)$

Girl = A

Pants = B

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

似然（类条件概率）
假设A为真的条件下，观察到证据B的概率

（后验概率）
在证据B出现后，假设A为真的概率

先验概率
样本空间中各类样本所占的比例，可通过各类样本出现的频率估计（大数定理）

边缘概率
观察到证据B的总概率

1、朴素贝叶斯分类器采用了"属性条件独立性假设"

即每个属性独立地对分类结果发生影响。

2、西瓜分类---------重点计算

➢ 西瓜分类

现在有一个新西瓜的样本：
青绿，稍蜷，浊响，清晰

使用朴素贝叶斯算法计算
这个新西瓜是好瓜的概率

编号	色泽	根蒂	敲声	纹理	好瓜
1	青绿	蜷缩	浊响	清晰	是
2	乌黑	蜷缩	沉闷	清晰	是
3	乌黑	蜷缩	浊响	清晰	是
4	青绿	蜷缩	沉闷	清晰	是
5	浅白	蜷缩	浊响	清晰	是
6	青绿	稍蜷	浊响	清晰	是
7	乌黑	稍蜷	浊响	稍糊	是
8	乌黑	稍蜷	浊响	清晰	是
9	乌黑	稍蜷	沉闷	稍糊	否
10	青绿	硬挺	清脆	清晰	否
11	浅白	硬挺	清脆	模糊	否
12	浅白	蜷缩	浊响	模糊	否
13	青绿	稍蜷	浊响	稍糊	否
14	浅白	稍蜷	沉闷	稍糊	否
15	乌黑	稍蜷	浊响	清晰	否
16	浅白	蜷缩	浊响	模糊	否
17	青绿	蜷缩	沉闷	稍糊	否

$P(h|x) = \frac{P(h) \cdot P(x|h)}{P(x)}$

$\propto P(色=青|h^+) \cdot P(根=蜷|h^+) \cdot \cdots$

第七章 聚类----计算重点

1、聚类在“无监督学习”任务中研究最多、应用最广。

2、聚类相似性度量

常用距离形式 **闵可夫斯基距离**

➢ 常用距离形式

闵可夫斯基距离（Minkowski distiance）

闵可夫斯基距离是曼哈顿距离和欧氏距离的扩展，定义为：

$$d(\boldsymbol{x},\boldsymbol{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}$$

其中，p是一个可变参数。可以发现：

当$p = 2$时，闵可夫斯基距离即为欧氏距离；

当$p = 1$时，闵可夫斯基距离则为曼哈顿距离。

3、其他度量形式：-------重点计算

汉明距离：

➢ 其它度量形式：

汉明距离（Hamming Distance）

- 给定两个相同长度的字符串，汉明距离定义为两个字符串中对应位置数值不同的字符的个数
- 如字符串“101101”和“100110”之间的汉明距离为3

汉明距离越小，两个样本特征越相似

思考H(x,y) =? For x=[0, 1, 0, 1, 1, 0]t;
y=[0, 0, 1, 1, 0, 1]t;

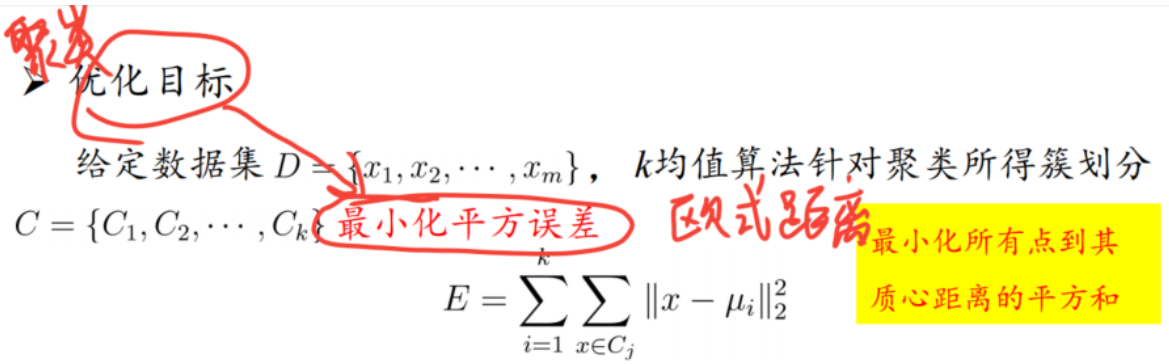
4、原型聚类----考计算

主要介绍K-means算法

1、初始化质心

2、判断所有样本到各个质心距离

5、K-means距离的度量：常用（欧几里得）距离和余弦相似度



其中，μ_i是簇 C_i 的均值向量

E值在一定程度上刻画了簇内样本围绕簇均值向量的紧密程度，E值越小，则簇内样本相似度越高。

算法流程（迭代优化）：
初始化每个簇的均值向量
repeat
1. （更新）簇划分；
2. 计算每个簇的均值向量
until 当前均值向量均未更新

第十一章 卷积神经网络----重点计算

1、卷积核全连接由什么关系区别?

卷积层有参数/权重

全连接层没有!

➢ CNN的架构 – 卷积 （Convolution）

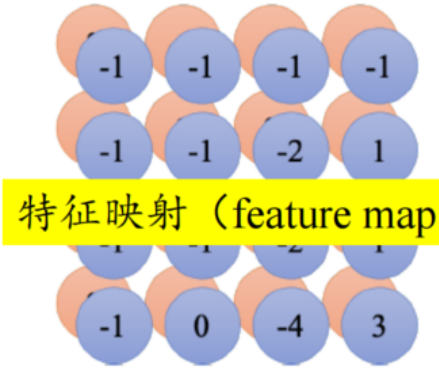
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 × 6 图像

-1	1	-1
-1	1	-1
-1	1	-1

滤波器 2

对每个滤波器做同样的过程



特征映射（feature map）

卷积神经网络

➢ CNN的架构 – 汇聚 （最大池化，Max Pooling）

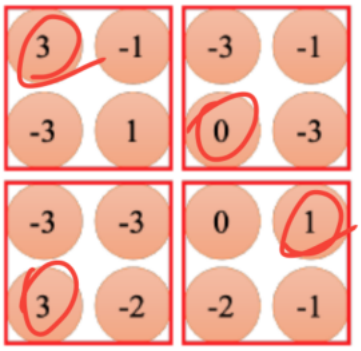
对卷积结果进行2x2无重叠最大池化操作

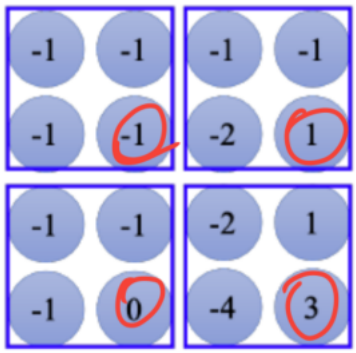
1	-1	-1
-1	1	-1
-1	-1	1

滤波器 1

-1	1	-1
-1	1	-1
-1	1	-1

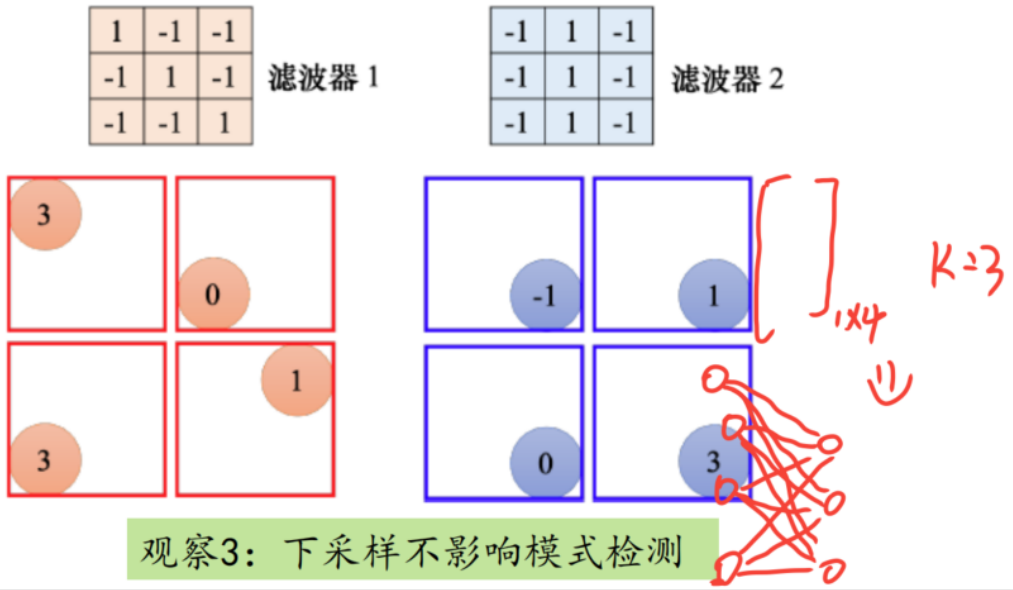
滤波器 2





➢ CNN的架构 – 汇聚 （最大池化，Max Pooling）

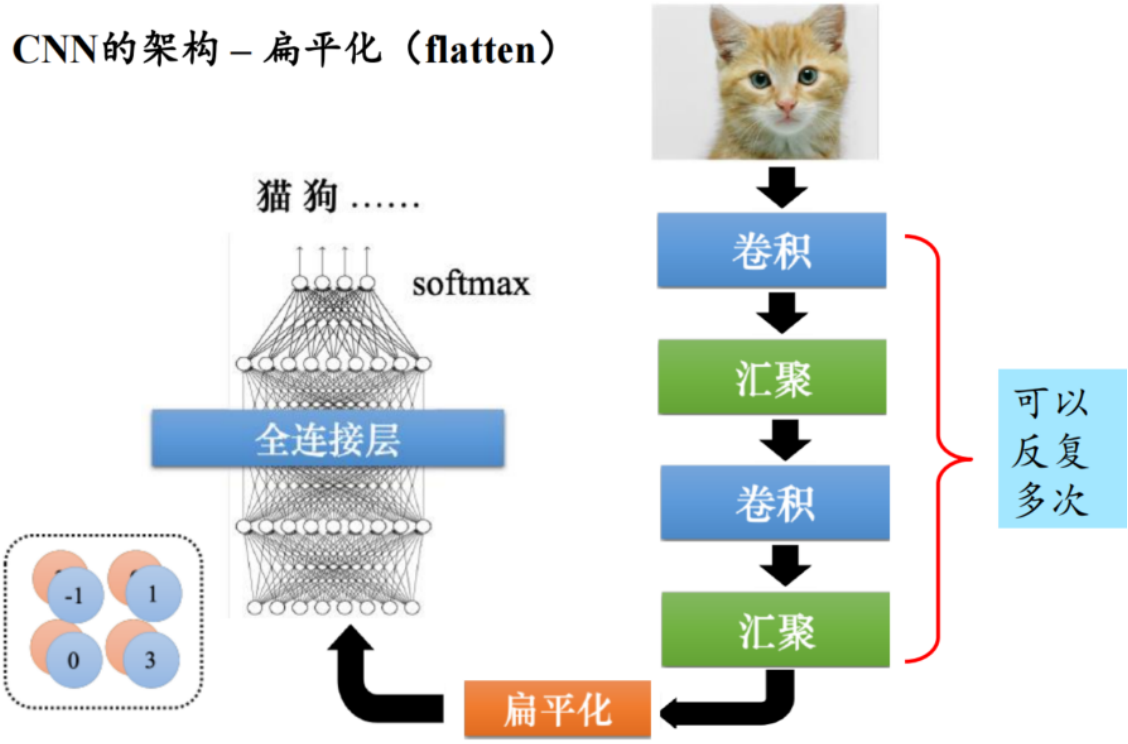
对卷积结果进行2x2无重叠最大池化操作

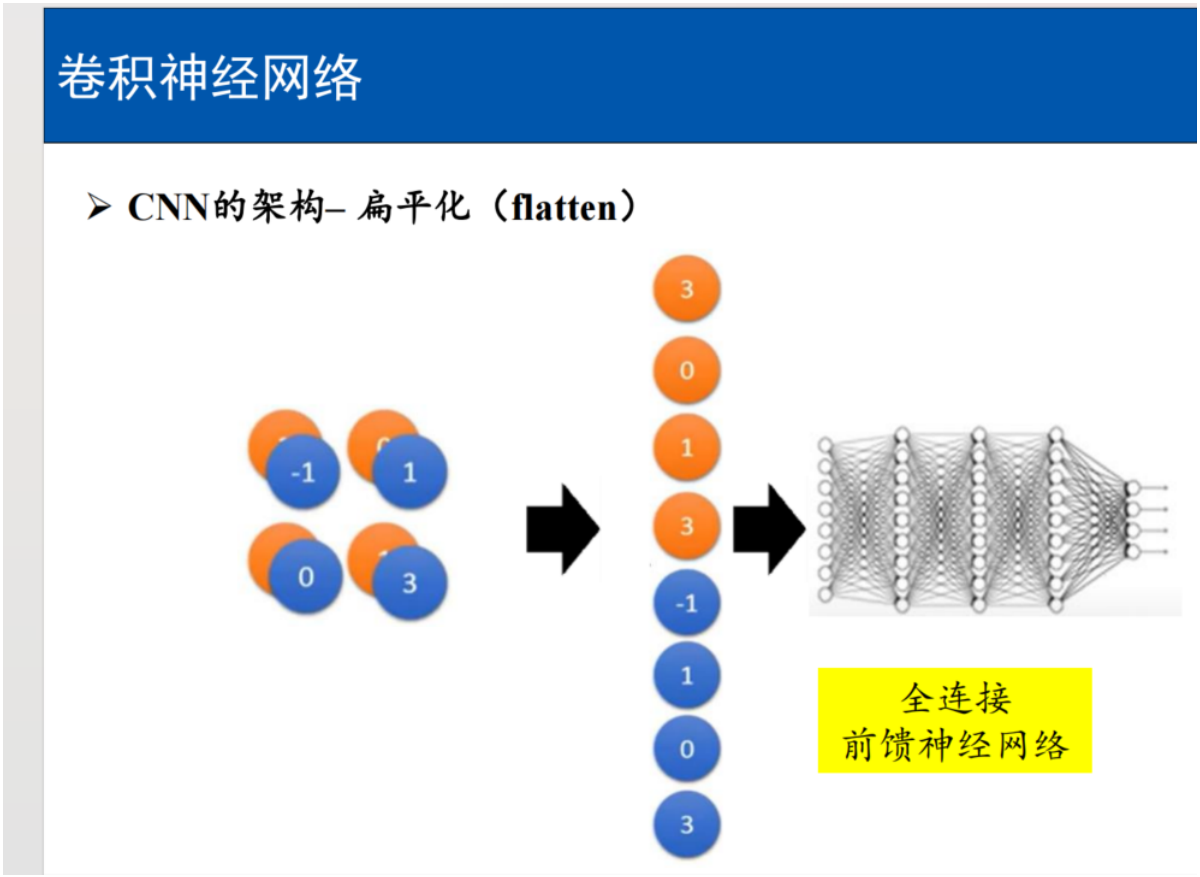


观察3：下采样不影响模式检测

卷积神经网络

➢ CNN的架构 – 扁平化（flatten）





二、病毒实验提交-----周五23：59

三、周四网络服务与配置--中午

四、网络服务与配置作业---今晚写上