1. 论述什么是多媒体？多媒体分析与理解有哪些应用领域和挑战？ （10 分）
2. 多媒体：多媒体是使用不同内容形式的组合的内容，如文本、音频、图像、动画、视频和交互式内容。
3. 应用领域：应用广泛，数据量大，就数字图像视频而言，就广泛应用在安防、教育、通讯、娱乐等产业中，多媒体还应用在图片检索、内容推荐、移动通迅、社交媒体平台、视频网站等。
4. 面对挑战：

* 如何表示多媒体数据（海量、高维、非结构化）
* 如何理解多媒体数据（语义鸿沟）
* 如何挖掘多媒体数据之间的相互关联（关联复杂，关联的协同性与互补性）
* 如何满足用户多样化的信息需求（海量数据、用户偏好与个性化）

1. 针对文本、音频与图像数据的特征表示方法，请分别列举 2 种典型特征，并分析其优缺点。（15 分）
2. 文本

* 词频（TF）表示法

优点：用词在文档中的出现频率可表示出一篇文档中的侧重点，便于统计与分析。

缺点：文中多次出现的介词、系动词等也会被统计出较高的权重。

* 潜在语义分析法（LSA）

优点：通过降维，有效解决了一词多义与一义多词的问题。

缺点：仍未解决文档中间的有序性被破坏的问题。

1. 音频

* 过零率

优点：能侧面反映信号在短时帧内的平均频率。

缺点：只关注了短时窗内的幅值信息，频域信息缺失。

* 梅尔频率倒谱系数（MFCC）

优点：对特征进行了去相关和压缩。

缺点：对所有频率段信息进行同等处理，不能突出重要信息。

1. 图像

* LBP

优点：一定程度上消除了光照变化的问题，具有旋转不变性，计算速度快。

缺点：光照不均时对应的LBP算子会发生变化，同时LBP也丢失了方向信息。

* SIFT

优点：具有很好的尺度不变性和鲁棒性。

缺点：实时性不高，具有时特征点较少，对边缘光滑的目标无法准确提取特征点。

1. 卷积神经网络中一个典型层通常包含三种基本操作，请回答每种操作的基本含义或类型，所具备的基本特性或者优缺点。（15 分）

典型层三种基本操作：卷积—非线性变换—池化

1. 卷积操作：

* 基本含义或类型：卷积操作又称滤波，卷积核函数又称滤波器。在输入图像上的二维卷积，卷积输出称为feature map。
* 基本特征或优缺点：通常会在同一个卷积层使用多个不同的卷积核以学习图像的不同特征。当卷积核输入包含多个通道时，可以把卷积核看成是3D的。

1. 非线性变换：

* 基本含义或类型：先通过Φ(x)将x空间的点转换成z空间的点，而在z空间上得到一个线性的假设，再恢复到原来的x空间中得到一个二次的假设。
* 基本特征或优缺点：优点是有很好的数学性质，缺点是很容易饱和、输出不是0均值的会对梯度产生影响。

1. 池化操作：

* 基本含义或类型：池化函数使用某一位置的相邻位置的总体统计特征来代替网络在该位置的输出。常用的池化函数有最大池化、平均池化。
* 基本特征或优缺点：当输出作出少量平移时，池化能够帮助输入的表示近似不变,即平移不变。减小参数规模，提高统计效率。在全连接层之前采用全局池化，可以保持全连接层节点数量不变，不受输入图像大小的影响。

1. 请说明图像语义理解的基本研究内容，及其各项内容的含义。请针对其中任一项内容，从中选取 1 个典型算法或模型对其具体实现过程进行详细说明。（15 分）
2. 图像语义理解本质上是学习图像底层特征与高层语义的映射关系，研究图像中有什么目标、目标之间的相互关系、图像是什么场景的一门学科。主要研究图像分类、图像标注、图像目标检测、图像语义分割、图像语义描述、视觉问答等内容。
3. 图像分类：给每幅图像一个类别标签（根据各自在图像信息中所反映的不同特征，把不同类别的目标区分开来的图像处理方法。它利用计算机对图像进行定量分析，把图像或图像中的每个像元或区域划归为若干个类别中的某一种，以代替人的视觉判读。）

图像标注：给图像多个语义标签

图像目标检测：对图像中的目标标注一个类别同时用一个紧致框定位目标（输出一系列带有类别标签的目标框）

图像语义分割：给图像多个语义标签且像素级定位

图像语义描述：用自然语言来描述图像内容，即看图说话。

视觉问答：给定图片和图片相关的问题，回答出正确的答案。

1. 基于深度学习的目标检测中的R-CNN模型

* 首先输入测试图像；
* 利用selective search算法在图像中提取2000个左右的目标候选区域；
* 将每个候选区域缩放成227x227的大小输入到CNN，采用fc7层特征输出；
* 基于CNN的候选区域特征表示，输入到SVM进行分类。

1. 图像语义理解面临的主要困难

* 计算机依据图像视觉底层进行相似性判断，人类根据语义信息对图像进行相似性判断，两者之间产生“语义鸿沟”。
* 图像数据的多样性与复杂性。多样性和复杂性包括：光照条件、尺度变化、遮挡模糊、类间差距大，类内差异小。

1. 简述BP算法的基本原理，并说明该算法在神经网络中可能遇到的问题及应对措施？
2. 基本原理

利用输出后的误差来估计输出层的前一层的误差，再用这个误差估计更前一层的误差，依此逐层反传下去，从而获得所有其它各层的误差估计。

1. 可能遇到的问题与解决措施

* 梯度消失与梯度爆炸。

解决措施：权值更新过程与激活函数的导数有关，随着训练的进行，激活函数进入饱和区，梯度调节几乎停止，解决措施是利用ReLU代替Sigmoid，或者使用批正则化，避免网络的梯度消失，提高模型的泛化能力；为防止梯度爆炸，可通过设置梯度剪切阈值的方法。

* 非凸问题求解，容易陷入局部极小值。

解决措施：采用参数初始化或者预训练模型，例如采用标准初始化、He初始化等；自适应调节学习率；引入动量的梯度下降。

* 模型过拟合

解决措施：在适当时停止训练；扩增训练样本；采用正则化的方法（岭回归、LASSO回归）；采用Dropout技术。

1. 简述特征表示学习的必要性，及其主要解决方法和基本原理。
2. 主要解决方法和基本原理

* 特征降维
* PCA：通过主成分分析方法，将高维数据映射到低维空间，并期望在所投影维度上数据的方差最大，以此使用较少的数据维度，同时保留住较多的原数据点的特性。
* LDA：主要针对两类问题，将数据有原始维度投影到一维空间，可直接用于分类。降维投影后的样本点，同类样本之间距离尽可能近，异类样本之间距离尽可能远。
* 流形学习（LLE、Isomap）：从高维采样数据中恢复低维流形结构，即找到高维空间中的低维流形，并求出相应的嵌入映射，以实现维数约简或者数据可视化。
* 特征选择
* 过滤法：通过分析特征子集有效性来衡量其好坏，一般用作预处理，与模型学习无关。
* 包裹法：实质上是一个分类器，用选取的特征子集对样本集进行分类，模型性能作为衡量特征子集好坏的标准。
* 嵌入法：在模型学习的同时，学习出对提高模型性能最好的特征子集，如L1/L2正则化。
* 特征融合
* 前向融合：特征层融合，形成融合特征表示，然后进行模型的学习。
* 后向融合：决策层融合，不同特征分别进行模型学习后融合，得到最后的融合分类结果。

1. 详细说明基于 SVD 和 RBM （这两种算法是NetFlix比赛中单模型性能最好的两种方法）推荐方法的基本原理， 并比较它们的优劣。（15 分）
2. SVD推荐方法的基本原理：对于所有用户对所有商品打分，可以建一个矩阵来表示，而这个矩阵时极其稀疏的。而对于任意矩阵都有它的满秩分解。

RU×X=PU×K·QK×I（U为用户数，I为商品数）

然后用R中已知数据去训练P和Q，使得P和Q相乘的结果最好的拟合已知的评分。

预测用户U对商品的评分：

假设已知的评分r，则预测误差为

继而可以计算出总的误差平方和：

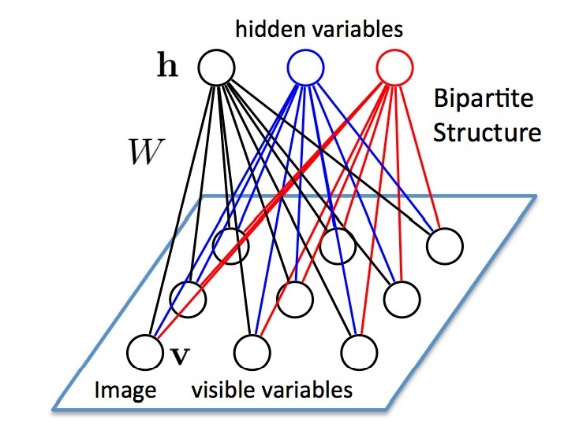


接下来训练是的SSE最小的模型即可。

1. RBM推荐方法的基本原理：

受限玻尔兹曼机简单介绍

传统的受限玻尔兹曼机是一种如下图所示, 其由一些可见单元(visible unit，对应可见变量，亦即数据样本)和一些隐藏单元(hidden unit，对应隐藏变量)构成，可见变量和隐藏变量都是二元变量，亦即其状态取{0,1}。整个网络是一个二部图，只有可见单元和隐藏单元之间才会存在边，可见单元之间以及隐藏单元之间都不会有边连接。



将该模型应用到协同过滤需要解决以下两个问题:

鉴于RBM中的单元都是二元变量, 如果用这些二元变量来对整数值的评分建模?

用户的打分是非常稀疏的, 亦即用户只会对很少的物品(比如电影)打分, 如何处理这些缺失的评分?

2. 基于RBM的协同过滤

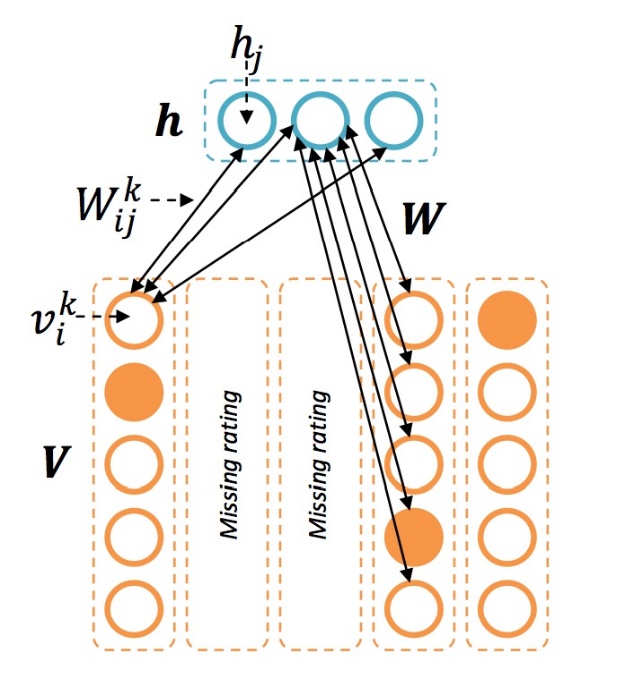
R. R. Salakhutdinov等人提出了一种使用RBM来进行协同过滤的方法:

假设有m个电影, 则使用m个softmax单元来作为可见单元来构造RBM.  对于每个用户使用不同的RBM, 这些不同的RBM仅仅是可见单元不同, 因为不同的用户会对不同的电影打分, 所有的这些RBM的可见单元共用相同的偏置以及和隐藏单元的连接权重W. 该方法很好的解决了之前提到的问题:

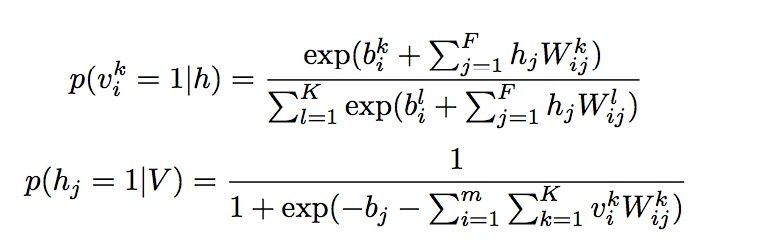
使用softmax来对用户的评分进行建模, softmax是一种组合可见单元, 包含k个二元单元, 第i个二元单元当且只当用户对该电影打分为i时才会置为1.

如果一个用户没有对第j个电影评分, 则该用户的RBM中不存在第j个softmax单元.

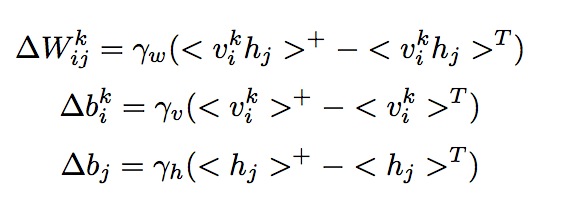
该模型如下图所示:



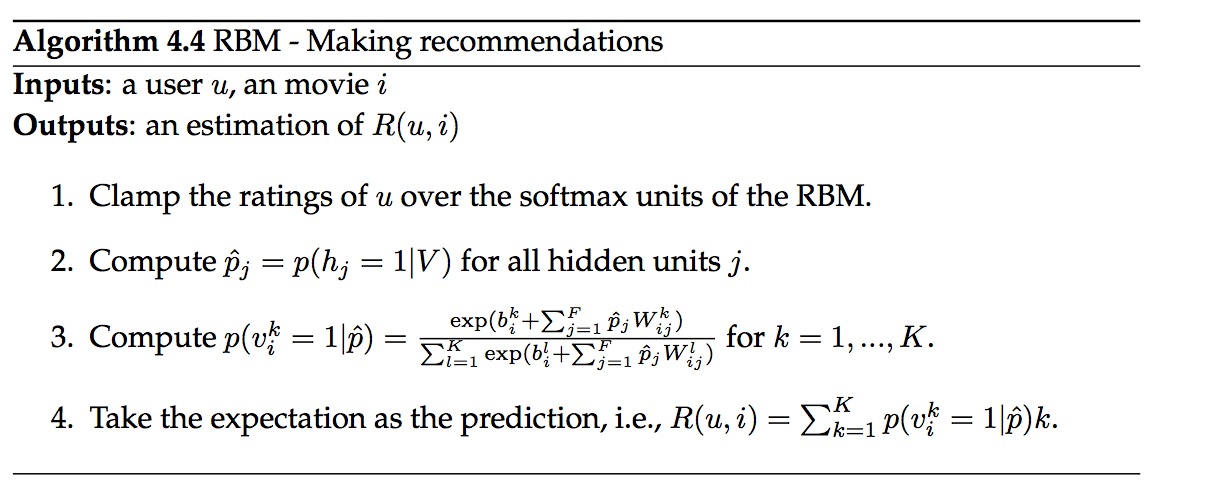
可是单元V和隐藏单元h的条件概率为:



模型参数的学习过程非常类似于RBM的DC算法:



训练完模型后, 计算用户对未评价物品的预测评分的算法为:



将某一用户对某一商品的评分视作一个softmax神经元，softmax神经元是个长度为k的向量，只有一个分量为1，其余分量为0。而未评分的 部分就可以用全0的softmax神经元表示。这样某个用户的评分可以用矩阵V来表示，在给定可见单元的状态下其激活概率为：



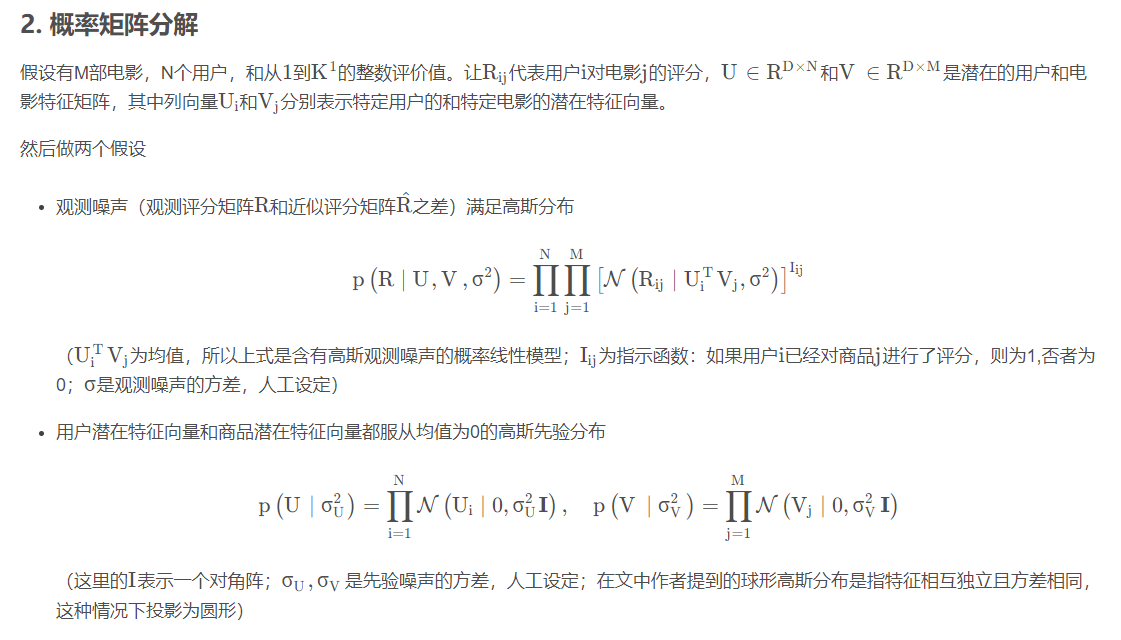
同理，在给定隐单元状态下 ，可见单元的激活率为：

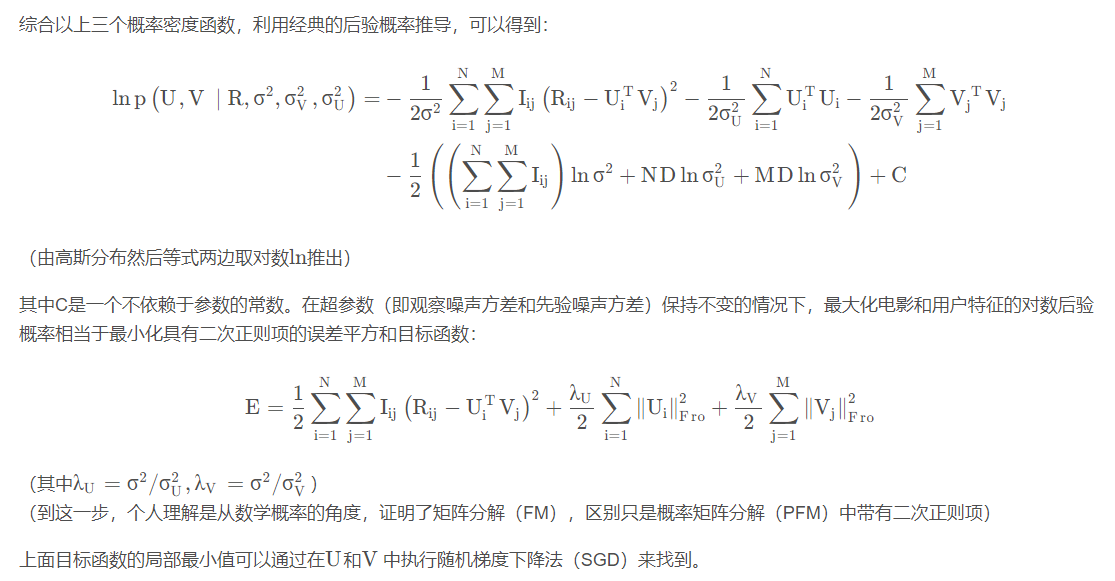


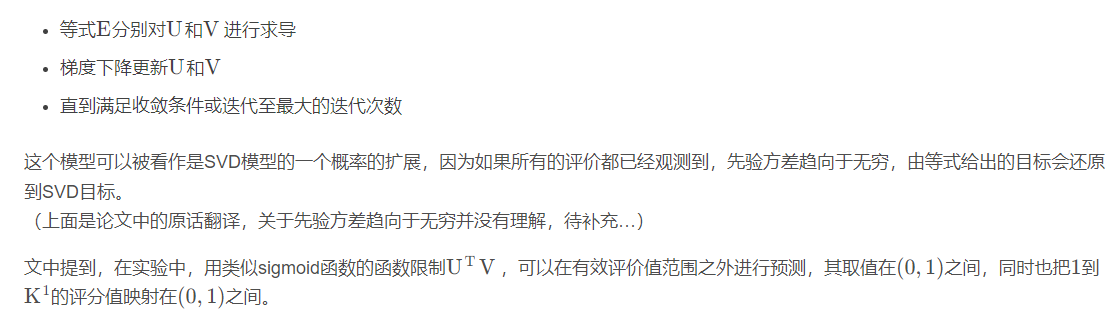
算每一个评分K所对应的激活概率，取所有概率的期望作为预测值。

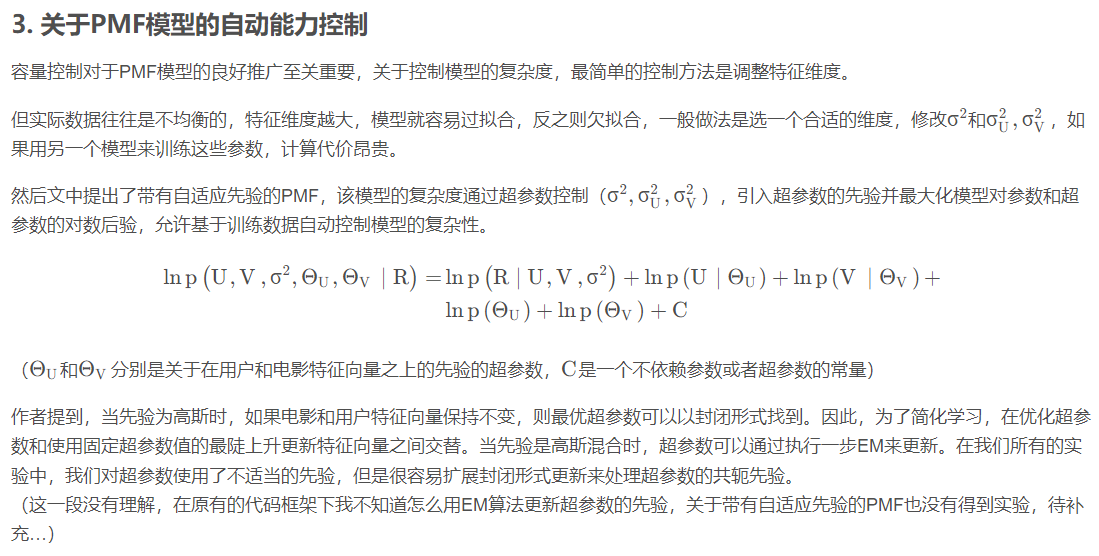
1. 比较：SVD计算过程虽更简单，但由于训练目标单一，容易造成过拟合，而RBM则可以防止梯度爆炸和梯度消失，但求期望的过程会比较复杂，学习效率过慢。
2. PMF

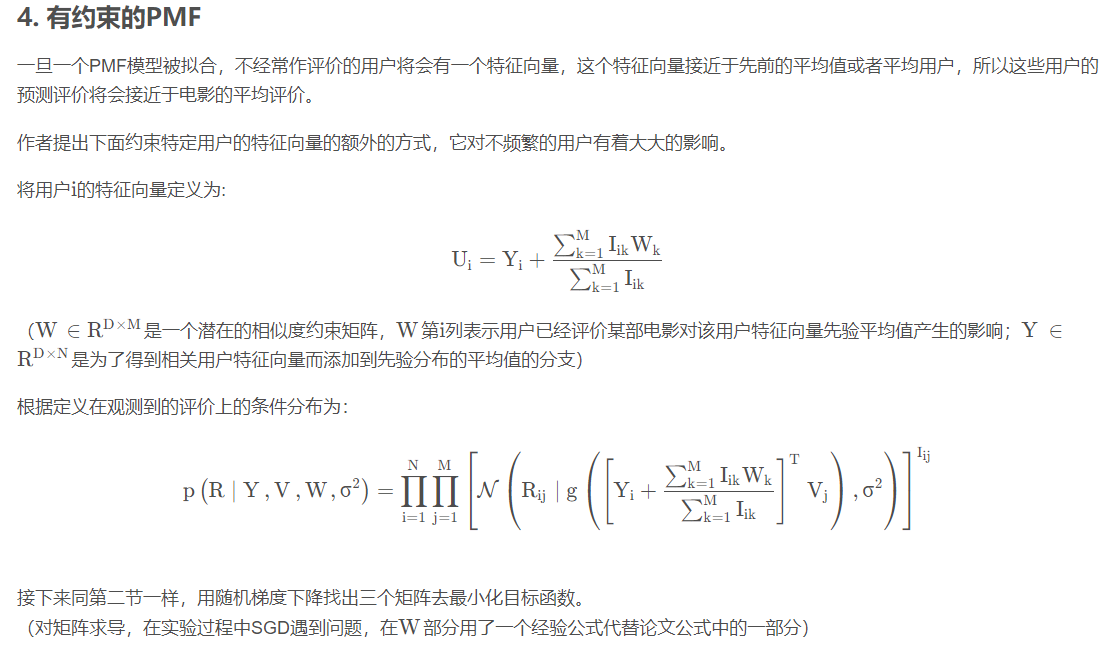










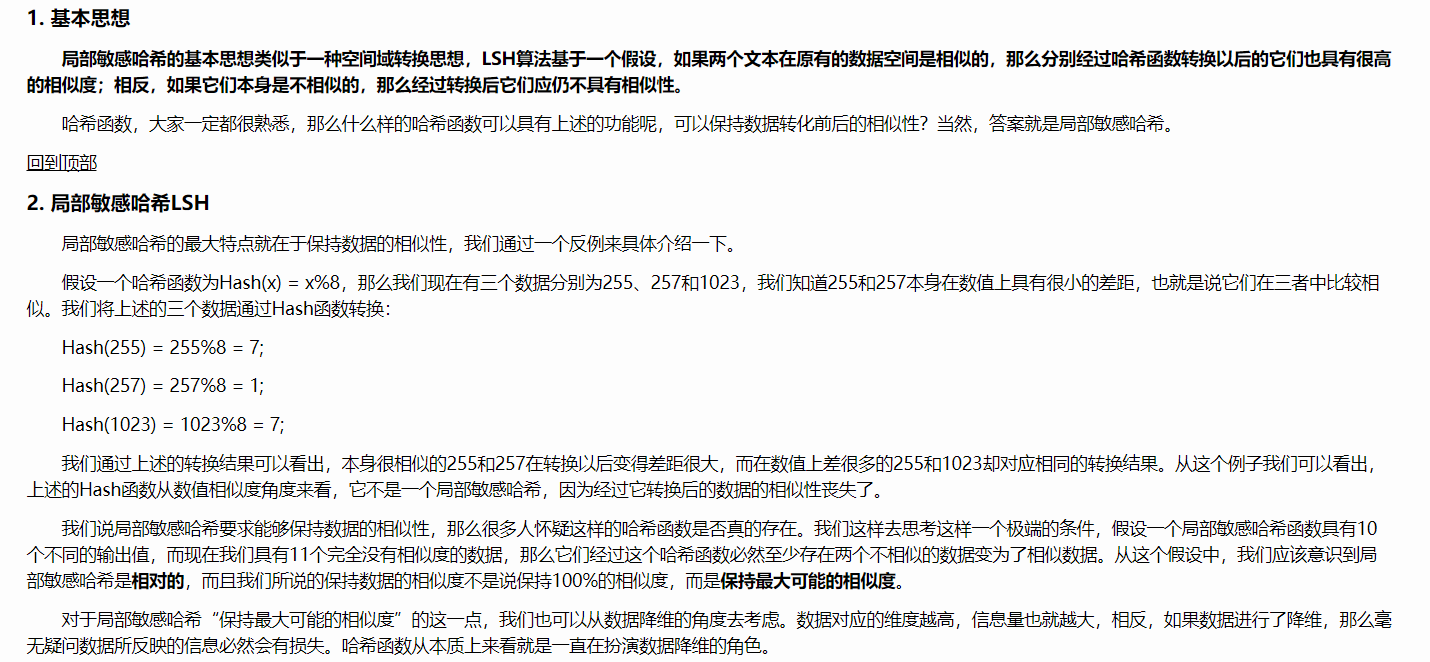


1. 简单描述迭代量化哈希方法(Iterative Quantization，ITQ)的基本思想,并比较 ITQ 方法与局部敏感哈希(LSH)方法的优劣。（15 分）
2. 迭代量化哈希方法的基本思想：先对数据集进行PCA降维，然后寻找量化误差最小的旋转矩阵即可得到对应该最优旋转矩阵下的特征向量的二进制编码。（查找 PCA 嵌入式数据；对于c\_bit代码，采用顶部cPCA方向和零阈值）
3. 比较ITQ方法与局部敏感哈希(LSH)方法的优劣：

* ITQ

优点：相比局部敏感哈希(LSH)方法多了一步操作，即数据降维后使用矩阵旋转优化，可以降低量化误差

劣势：由于PCA不同维度的方差不平衡，旋转PCA投影数据以尽量减少量化误差时需不断控制旋转角度，即找出最优的旋转矩阵和与之对应的编码，相对麻烦

* LSH
* 

优点：通过hash function映射变换操作，将原始数据集合分成了多个子集合，而每个子集合中的数据间是相邻的且该子集合中的元素个数较小，因此将一个在超大集合内查找相邻元素的问题转化为了在一个很小的集合内查找相邻元素的问题，计算量下降了很多，改善近似检索的计算表现。

劣势：局部敏感哈希(LSH)方法并不能保证一定能够查找到与query data point最相邻的数据

1. 运动目标检测的困难有哪些？并简述目前常用方法的优缺点。（15 分）
2. 困难：光照的变化、动态背景、伪装、相机抖动、相机自动对焦、间断的物体运动、阴影。
3. 目前常用方法：帧间差方法、背景差分法、统计平均、高斯模型、非参数核密度概率估计、基于码本的运动目标检测、背景建模（ViBe、SuBSENSE）

优缺点：

* 帧间差方法

优点：算法简单、容易实现、检测速度较快。而且一般相邻两帧的时间间隔比较短，因此对场景的光线变化不是特比敏感。

缺点：对噪声非常敏感且检测出的物体位置不精确，其次帧差法的检测结果与目标运动速度和相邻两帧间隔大小有关。目标运动过快会被分成两个目标；目运动过慢会被视作背景。

* 背景差分法

优点：算法比较简单；一定程度上克服了环境光线的影响；

缺点：不能用于运动的摄像头；对背景图像实时更新困难；

* 统计平均法

优点：选择适当的参数能够很好地修正背景图像，从而得到一个较为真实的背景估计图像。

缺点：对于频繁出现的运动目标或者在场景内停留的时间较长，该模型不能够很好的提取运动目标。在复杂场景中会将一些伪目标（如摇摆的树枝等）检测为运动目标。这是由于在场景中树枝的摇摆会引起像素值的变化。

* 混合高斯模型

优点：能适应背景随时间的缓慢变化，能描述背景中的如显示器屏幕闪烁，树枝晃动等一些周期性扰动等。

缺点：对于缓慢运动的目标不能精确的检测提取，不能够很好的适应阴影、噪声造成的误检或漏检，且不能适应场景的突然变化。

* 非参数核密度概率估计

优点：可以渐进收敛于任意形状的概率密度，而且对动态的场景还具有一定的适应性。

缺点：运算量非常大，很难实现对视频图像的实时检测。对内存要求比较高

* 基于码本的运动目标检测

优点：鲁棒性强，计算效率高，速度快，运算量少，精确度高。

缺点:当训练帧存在较大运动前景时，码本模型的建立会非常不准确，更新参数的调整非常复杂，不能广泛应用于实际情况；由于对于视频每个像素都要建立一个甚至多个码本模型，训练模型比较耗时，若需要背景重构，将大大降低运算速度。

* ViBe

优点：思想简单，易于实现。（用有限样本近似无限时间窗口）计算量小。运算效率高。（样本少；优化了相似度匹配算法）样本衰减策略。（随机更新策略使得样本生命周期指数衰减，不同于其他方法的先进先出）

缺点：有鬼影、静止目标、阴影前景和运动目标不完整等问题。

* SuBSENSE

优点：使用了反馈机制，更好适应不同场景，对噪声更为鲁棒。

1. 简述PageRank的基本原理。
2. PageRank的基本思想：如果一个网页被很多其他网页所链接，说明它受到普遍的承认和信赖，那么它的排名就越高。
3. PageRank的算法原理：

* 在初始阶段：网页通过链接关系构建起Web图，每个页面设置相同的PageRank值，通过若干轮的计算，会得到每个页面所获得的最终PageRank值。随着每一轮的计算进行，网页当前的PageRank值会不断得到更新。
* 在一轮中更新页面PageRank得分的计算方法为：



式中：

表示网页的后向链接

表示网页的前向链接数目

表示网页的PageRank

表示整个网络中网页的总数

表示阻尼系数。阻尼系数的引进原因：由于存在一些出链为0，也就是那些不链接任何其他网页的网，即孤立网页也使得很多网页能被访问到。