姓名：刘鑫杰 专业：计算机科学与技术

***The Seventh Week***

**前言：**

这周完成了LAD论文和代码的阅读和理解, 同时参考**《**Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers**》**看了关于ADMM算法原理. 同时也阅读了论文了《Joint Discriminative Attributes and Similarity Embeddings Modeling for Zero-Shot Recognition》, 并做如些的阅读报告.

**大目录：**

**一. LAD论文回顾及其源代码理解**

**二.ADMM算法原理**

**三.《联合判别属性与相似性零样本识别的嵌入建模》阅读报告**

**四.《matrix book》------2.5&2.6**

**一. LAD论文回顾及其源代码理解**

**1.LAD论文回顾**

LAD：在SAE的基础上对属性进行组合，学习隐含属性。利用字典学习框架将隐含属性空间与属性空间和相似性空间连接起来。通俗的说就是学习一个隐藏属性字典

补充: 字典学习🡪比如一个向量,是k维的,现在有n个k维向量, 即一个k \* n的字典, 其中n>>k, 所谓字典学习,就是将当前的一个k维向量, 通过该字典中各个向量的一个线性组合表示当前的这个k维向量.

**2.本论文提供解决问题的思路**

利用字典学习框架直接建模隐含属性空间，其中图像可以通过一些隐含属性的字典项重建。为了保留语义信息，利用线性变换来建立属性和隐含属性之间的关系，因此可以将隐含属性视为属性的不同组合。此外，为了使隐含属性具有辨别力，使用已见类分类器来对不同的类进行分类，其中概率输出可以看作与已见类的相似性。因此，我们可以将图像表示从隐含属性空间转换到相似空间。

**3.LAD算法流程和代码解读**

X是样本的特征表示

A是样本的语义表示

Z是样本类别标签

P是类别语义表示

Ys是重建系数

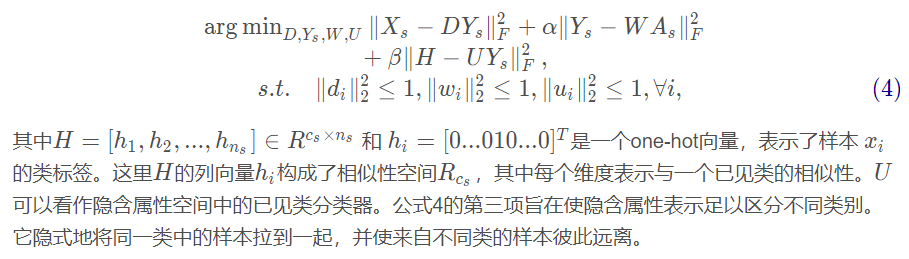
D是字典

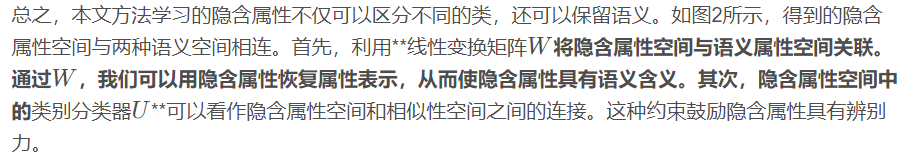
W线性变换矩阵

U一个从隐含属性空间到已见类的线性映射

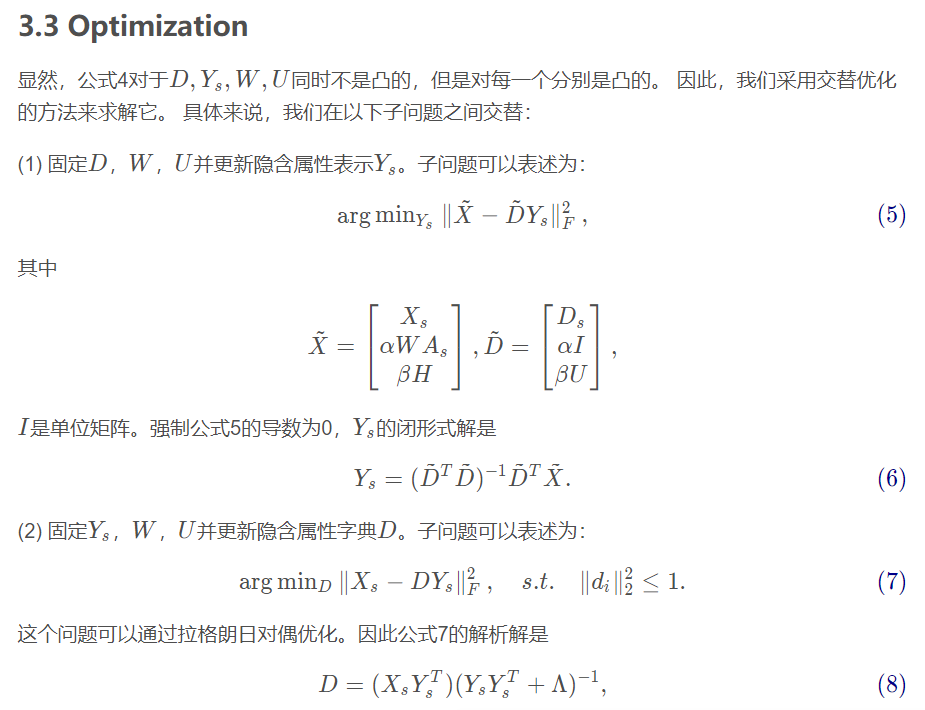
I是单位矩阵

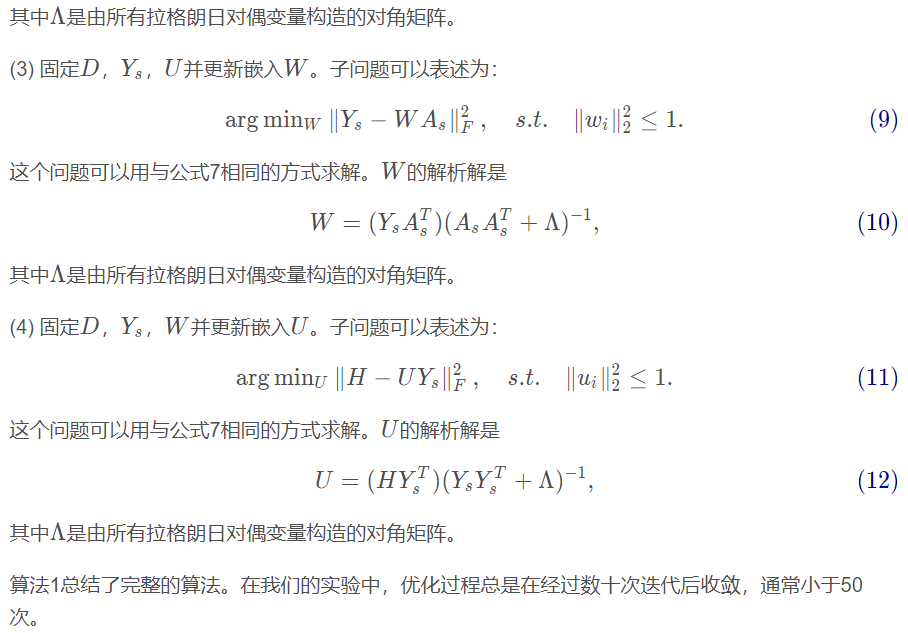
one-hot标签: 例如标签2对应的one-hot标签就是0 0 1 0; 使用one-hot 的直接原因是: 将类别哦变量转换为及其学习算法易于利用的一种形式的过程,

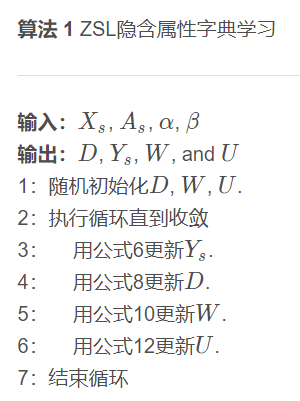




对上面的方程优化







参考文献:

<https://blog.csdn.net/cp_oldy/article/details/81900966>

**二.ADMM算法原理**

参考知乎: <https://www.zhihu.com/question/36566112>

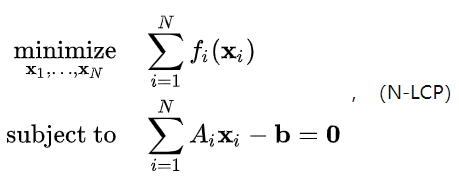
ADMM交替方向乘子法是机器学习中比较广泛使用约束问题的最优化方法. 适用于求解分布式凸优化问题: 即通过分解协调过程, 将大的全局问题分解为多个较小,较容易求解的局部子问题, 并通过协调子问题的解而得到大的全局问题的解

将无约束优化的部分用块坐标下降法来分别优化.

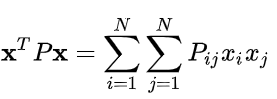
N-block ADMM的直接推广并不简单, ADMM最开始是针对2blocks的, 从2-blocks到N-block并不trivial(不是小事儿,不简单)

**N-block的直接推广:**

假设有一个线性约束的N-block凸优化问题



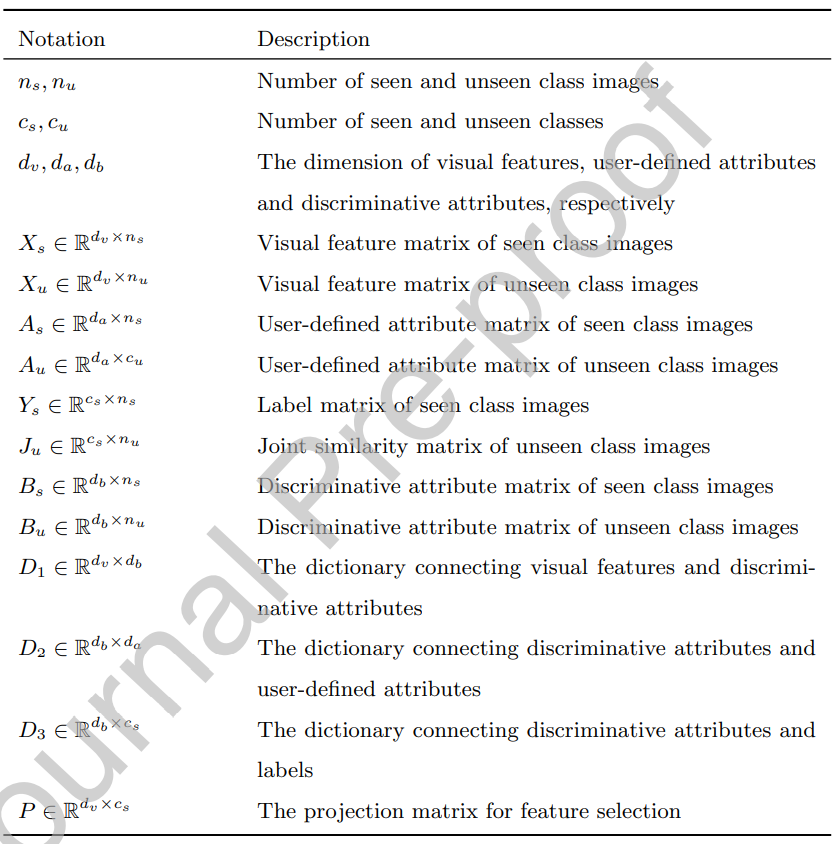
目标方程是由N个可分离的凸函数组成, 没有耦合项,如:



参考知乎

**三.《联合判别属性与相似性零样本识别的嵌入建模》阅读报告**

**3.1关键**:通过标签回归来学习有区别的属性,在语义相似度空间产生了更好的性能. 视觉相似性嵌入与语义相似度嵌入结合,丰富了联合相似度空间中的信息,大大缩小了视觉特征与语义属性之间的差距.

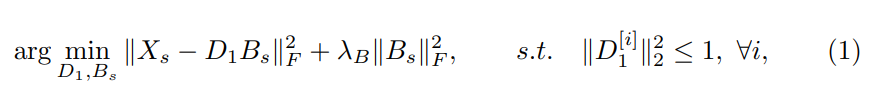


Ys的每一列表示一个one-hot编码标签向量 Cs是可见类别的数量

方法:

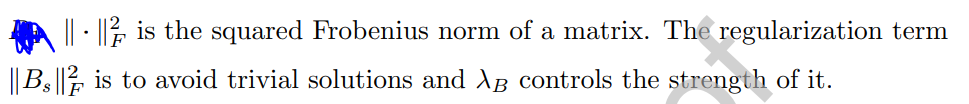
**step1:**

使用字典D1来构造视觉特征和判别属性之间的关系.



Bs是可见类图像的判别属性矩阵

db是判别属性的维度.



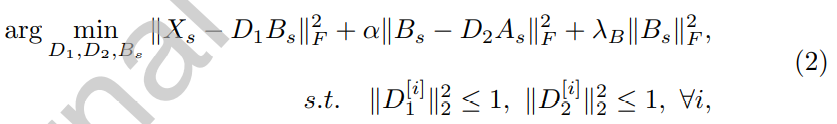
矩阵平方的F-范数🡪 F-范数:定义为矩阵A各项元素的绝对值平方的总和开根号

trivial solution平凡解,显然解: 显而易见的解,没有讨论的必要但为了结果的完整性仍需要考虑的结果.

λB控制它的强度

**step2:**

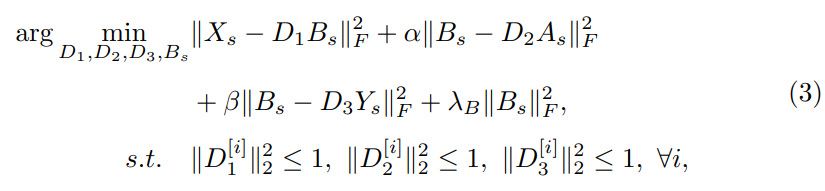
然后学习更多的区分属性, 这些区分属性是用户定义属性的线性组合🡪可以使用字典D2(联合B和A之间关系的矩阵)从用户定义的属性中学习判别属性Bs



alpha控制它的重要性的参数.

**step3**

至此, 已经利用用户定义的属性A来学习区分属性B, 并建立了视觉特征X和区分属性B之前的关系. 为了进一步利用标签监督, 我们引入了判别回归项来知道Bs的学习



D1: X和B D2: B和A D3: B和Y

beta是控制重要程度的参数

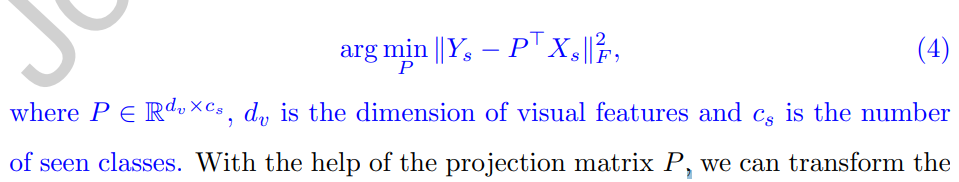
**3.2**视觉的相似嵌入

将用户定义的属性A和判别属性B转换为语义相似空间. 这个空间的维数与不可见类的个数相同, 每个维表示与相应类别的相似性.

基于以下事实提出视觉相似度嵌入:

(1)语义相似的图像通常在视觉上是相似的

基于此,构建视觉特征X和语义相似度之间的关系:

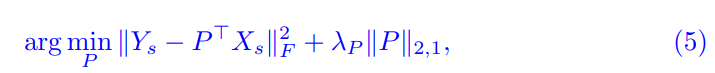


dv是视觉特征X的维度, Cs是可见类的数量

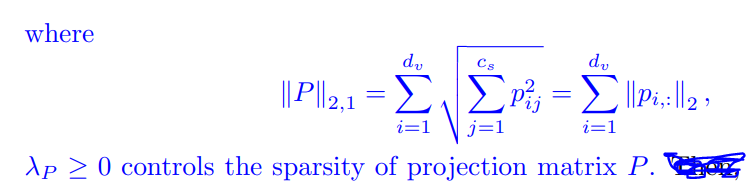
借助于投影矩阵P, 可以将不可见类的图像的视觉特征转化为相似空间,称为视觉相似空间.

(2)视觉特征对目标识别的重要性并不总是相同

基于此,在投影矩阵上加上一个2,1范数正则化, 以执行选择



其中

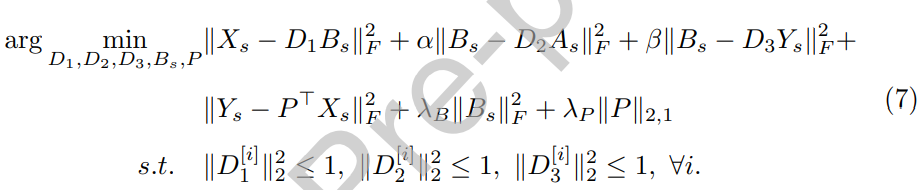


投影矩阵的2,1范数,是投影矩阵P的每一行元素的平方,求和,在开根号. 然后将每一行的结果相加. 这里的lamda\_p控制投影矩阵的稀疏性(非零元素多算稀疏)

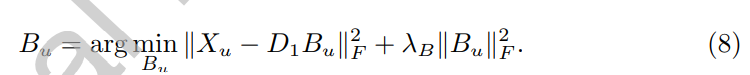
视觉相似性嵌入是通过将图像的视觉特征X通过投影矩阵P映射到视觉相似空间来实现的, 在视觉相似性空间中, 每个维度都表示与一个所见类别的相似性,因此可以通过最近邻方法对不可见类的图像进行识别.

**3.3** 联合判别属性与相似嵌入建模

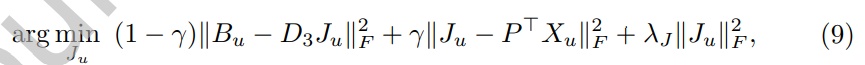
在训练阶段, 同时学习区分属性和视觉相似性嵌入, 训练的目标可以写成:



然后通过以下方法得到不可见类图像的判别属性



最后, 将不可见类的图像的区分属性Bu和视觉相似性嵌入合并到一个统一的框架中



Ju是不可见类的联合相似矩阵(它的每一列表示第i个测试图像和所看到的类之间的相似性).

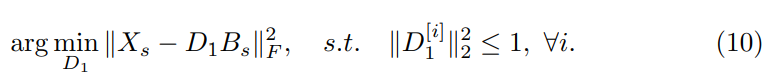
γ∈[0,1]为平衡参数。λJ控制正则化的强度。

**3.4** 优化

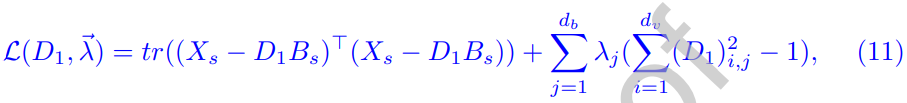
由于上述等式对于D1,D2,D3,Bs和P是非凸的.

使用ADMM算法来优化

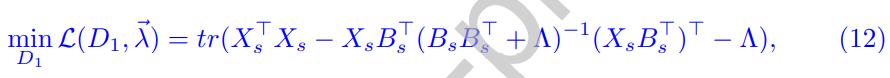
**step1:** 固定D2,D3和Bs来计算D1. 这样式(7)可以转化为具有二次约束的最小二乘问题

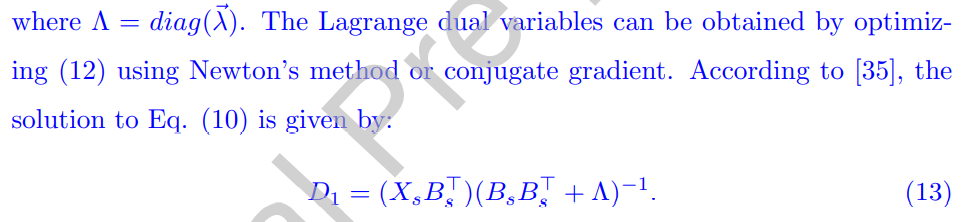


这个问题通常可以采用梯度下降法来解决,但是in this work 我们使用更有效的**拉格朗日对偶**来解决这个问题.



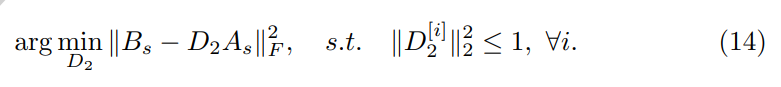
其中每个lamda\_j>=0是一个对偶变量. 然后, 得到了一下拉格朗日对偶化通过最小化D1解析

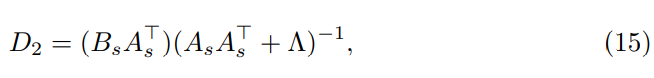




拉格朗日对偶变量可以通过牛顿法或者共轭梯度法优化得到.

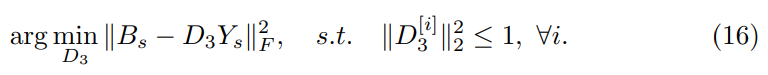
**step2:** 固定D1，D3，Bs来计算D2。式（7）可表述为

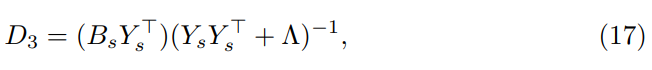


**🡪** 

式中，∧是一个对角矩阵，其对角元素为拉格朗日对偶变量，如等式（11）所示。

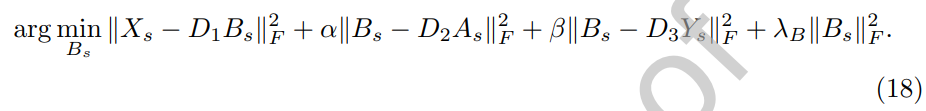
**step3:** 固定D1,D2,Bs来计算D3, 式（7）可表述为

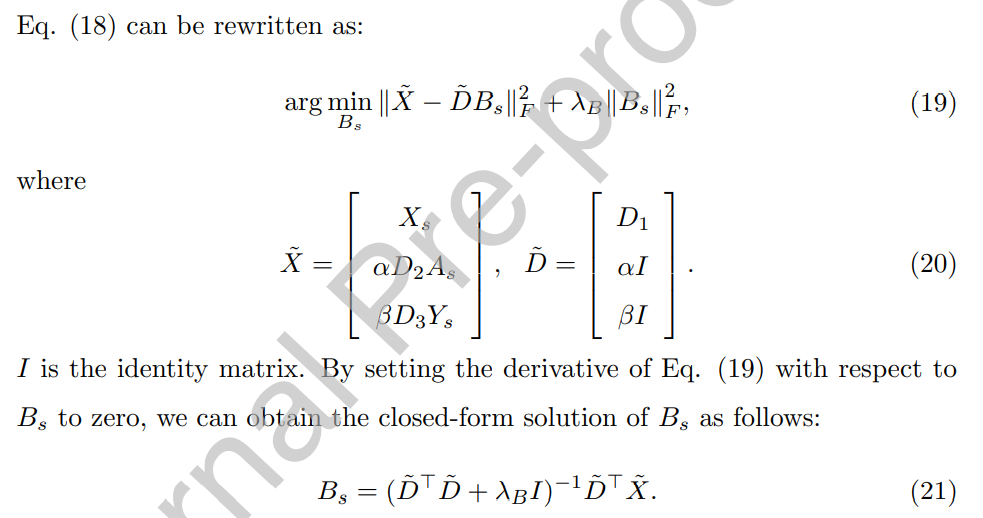


**🡪** 

式中，∧是一个对角矩阵，其对角元素为拉格朗日对偶变量，如等式（11）所示。

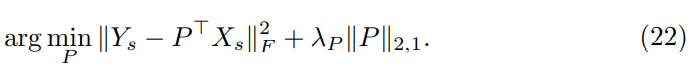
**step4:** 修正D1,D2,D3来计算判别属性Bs, 子问题可以表示为



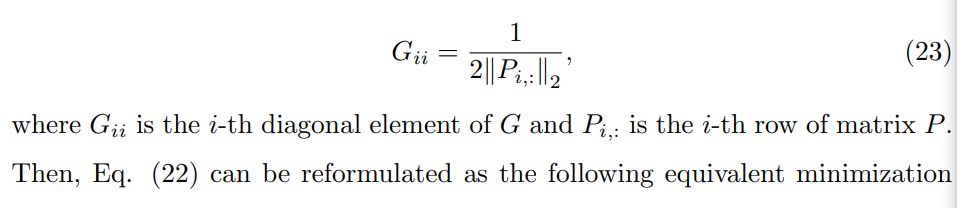


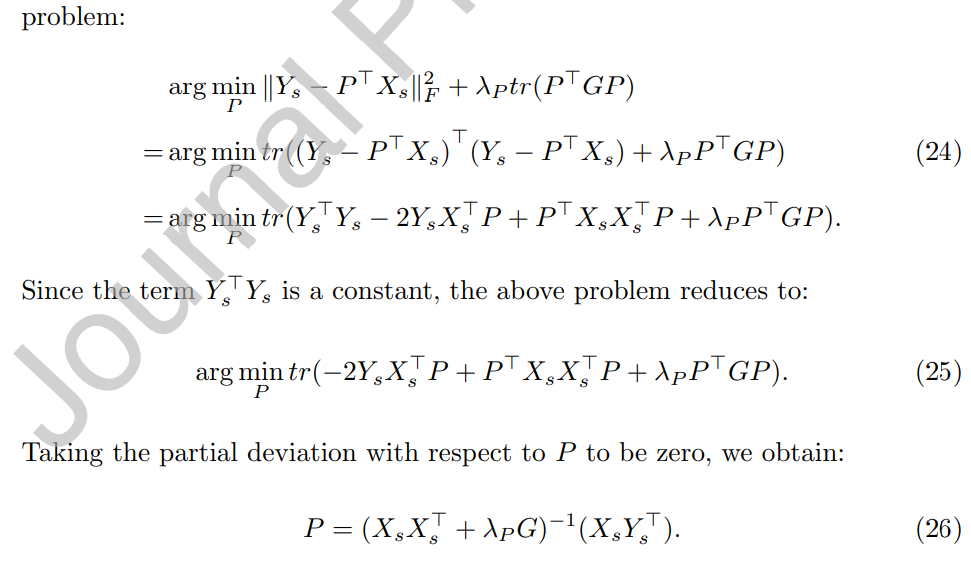
单位矩阵I, 通过将Bs的导数设为0, 可以得到Bs的闭式解

step5 计算最优P, 问题(7)可以归结为以下的优化问题:



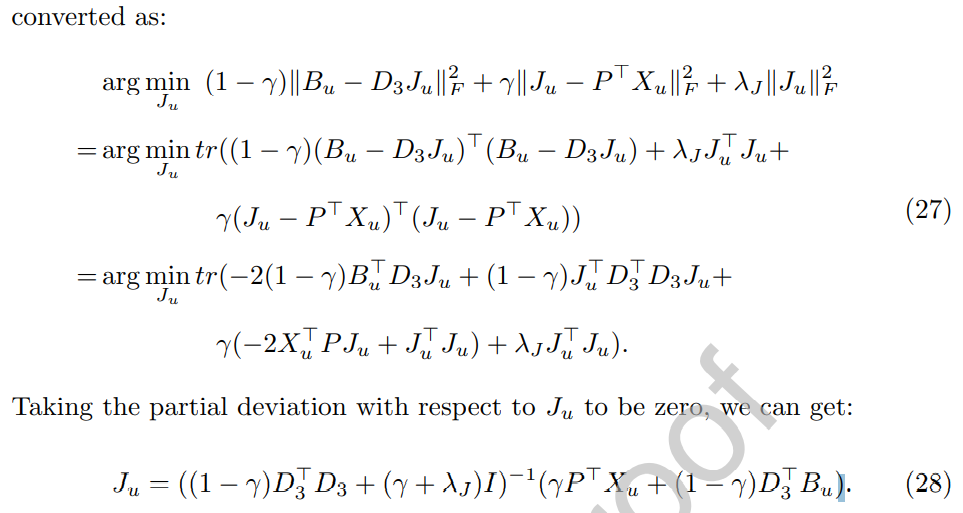
上述问题可以用2,1-范数极小化来解决。我们首先引入一个对角矩阵G，其定义如下：



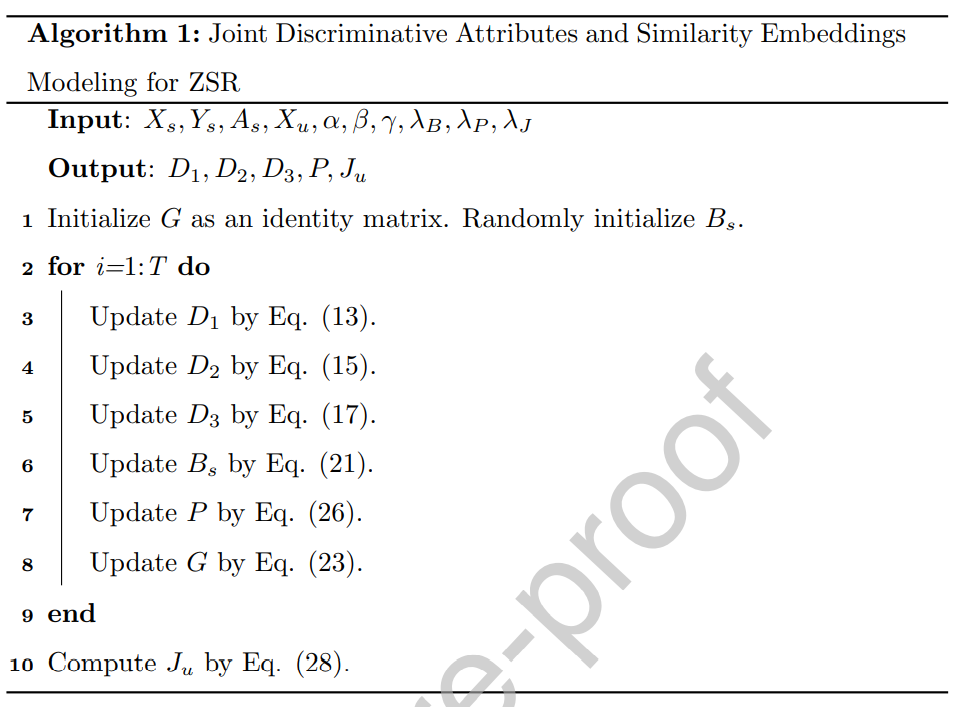


将P的部分偏差设为0, 得到(26)

**step6:**  计算最优的Ju, 优化问题(9)可以转化为:



完整的学习算法如下:



**3.5** 总结

在这个框架中有两个基本空间, 区分属性空间和联合相似空间,

1.区分属性空间:

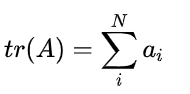
2.联合相似空间

difference:

该方法将语义相似度和视觉相似度结合起来，在联合相似度空间中对测试图像进行有效的分类

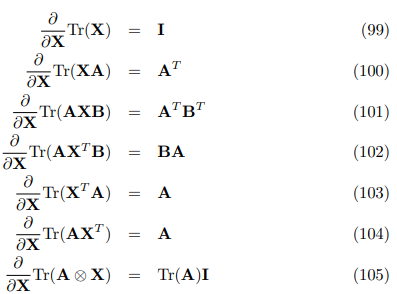
**四.《matrix book》------2.5&2.6**

4.1 矩阵迹的导数

矩阵迹：n x n矩阵A的主对角线上各个元素的和，记为

假设F（X）可微

一阶导数：



<https://www.cnblogs.com/qianxiayi/p/9025400.html>

<https://www.jianshu.com/p/9179ffda49a5>

<https://blog.csdn.net/u012421852/article/details/79594933>

矩阵的迹的意义：

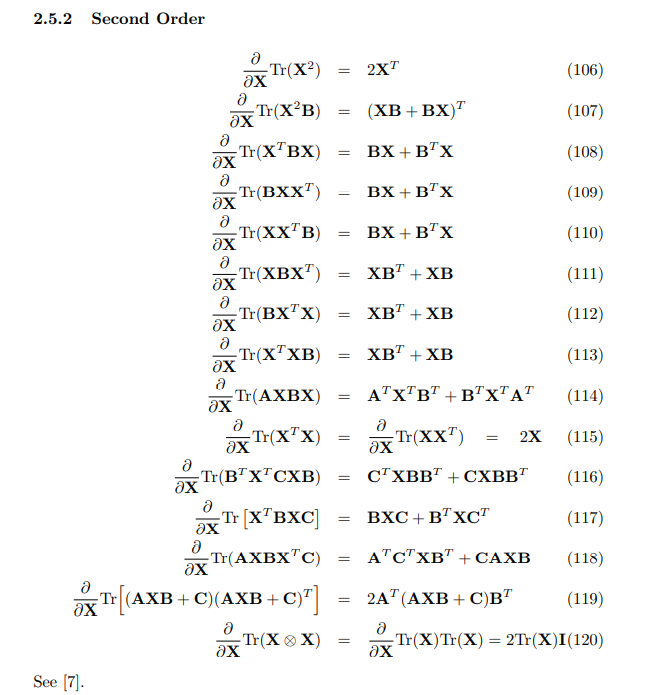
首先特征值的几何意义：在特征向量方向上的放缩（拉伸，压缩）的程度（或放缩的倍数）。

所以矩阵的迹的几何意义：所有线性无关的特征向量的放缩程度之和。迹=∑λ（特征值之和）。所以迹的几何意义与∑λ相同。

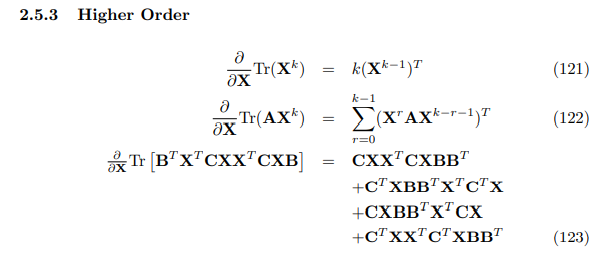
行列式的意义：体积

一个标量就是1x1的矩阵，迹就是它本身。行列式没有数模直观的几何意义，但迹的信息可以包含行列式。包含的意思是：如果你有一个矩阵A，Tr（A）并不能让你知道det(A). 但是如果你有关于迹的公式,即,你不但知道tr(A),还知道所有的tr(X),特别地,你就知道tr(A^2), tr(A^3),…, tr(A^n)这等于你知道A所有特征值的和,平方和,…,n次方和,这些信息加起来,可以推导出A所有特征值的积(det(A)) 反过来,如果知道det(X), 却并不能推出tr(A),所以迹是更加重要.

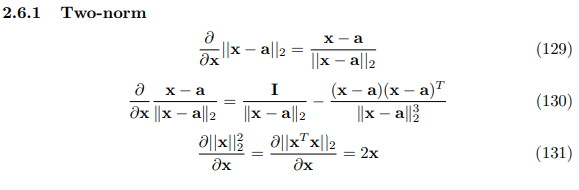
矩阵迹的二阶导数:



高阶导数



4.2向量范数的导数



向量范数的导数-->主要讲了矩阵A的二范数(空间上两个向量矩阵的直线距离,类似于求棋盘上的两点间的直线距离)的求导, 有三种形式.

范数:具有长度概念的函数.

