Assignment 1

161220096 欧阳鸿荣

一、问题

线性判别分析（LDA）和近邻成分分析（NCA）是两种在降维上被广泛使用的两种方法。请将他们同主成分分析（PCA）相比较并回答它们在数据规约上的基本原理**

二、解答

1.PCA的原理

PCA(Principal Component Analysis)是一种常用的数据分析方法。**PCA的原理是：基于线性变换（线性映射）的原理，将高维空间数据投影到低维空间上。**在数据分析上，实际上便是将数据的主成分（包含信息量大的维度）保留下来，忽略掉对数据描述不重要的成分。即将主成分维度组成的向量空间作为低维空间，将高维数据投影到这个空间上就完成了降维的工作。一般来说，PCA通过线性变换将原始数据变换为一组各维度线性无关的表示，在尽量减少数据损失的前提下提取数据的主要特征分量。

我们通过最大化方差法来推导PCA。假设我们将一个空间中的点投影到一个向量中去。对于原空间中的中心点，有：

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

假设u_1为投影向量，则投影之后的方差为：

$$\frac{1}{N}\sum_{n=1}^{N}(u_1^T x_n - u_1^T \bar{x})^2 = u_1^T S u_1$$

通过拉格朗日乘子法可以得到：

$$u_1^T S u_1 + \lambda_1(1 - u_1^T u_1)$$

对上式求导并令其为零，则可以得到：

$$S u_1 = \lambda_1 u_1$$

在上式中，λ对应特征值，u对应特征向量。假设要将一个D维的数据空间投影到M维的数据空间（M<D），那我们取前M个特征向量构成的投影矩阵就是能够使得方差最大的矩阵。

2.LDA的原理

LDA(Linear Discriminant Analysis)是一种**监督学习**的方法，**LDA的原理是将带上标签的数据（点）通过投影的方法投影到维度更低空间，使得投影后的点会按照类别区分为一簇一簇，使得相同类别的点在投影后的空间更加接近。**
LDA追求的目标是，给出一个标注了类别的数据集，投影到了一条直线之后，能够使得点尽量的按类别区分开。下面借由二分类简述下关于LDA的推导：

假设在二分类下的直线（投影函数为）：

$$y = w^T x$$

LDA分类的目标是一个目标是使得不同类别之间的距离越远越好，同一类别之中的距离越近越好。为了衡量这个因素，我们定义几个关键的值：

假设D_i表示属于类别i的的点，则有类别i的中心点为：

$$m_i = \frac{1}{n} \sum_{x \in D_i} x$$

类别i投影后的中心点为：

$$\widetilde{m_i} = w^T m_i$$

衡量类别i投影后，类别点之间的方差为：

$$\widetilde{s_i} = \sum_{y \in Y_i} (y - \widetilde{m_i})^2$$

从而可以得到LDA投影到w后的损失函数为：

$$J(w) = \frac{|\widetilde{m_1} - \widetilde{m_2}|^2}{\widetilde{s_1}^2 + \widetilde{s_2}^2}$$

我们定义一个投影前的各类别分散程度的矩阵

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T$$

则可以将损失函数化为：

$$J(w) = \frac{w^T S_B w}{w^T S_w w}$$

从而通过拉格朗日子乘法，并将分母限制为长度为1，则可以得到：

$$S_B w = \lambda S_w w$$

求出第i大的特征向量，就是对应的w_i了。

3.NCA的原理

NCA(Neighborhood Component Analysis)是一种与KNN相关联的距离测度学习算法，算法一般的应用场景是在原数据集上进行NCA距离测度学习，在这个过程中完成降维。该算法随机选择近邻，通过优化留一法的交叉检验结果来求得马氏距离中的变换矩阵。

假定n个输入样本$x_1, x_2, \ldots, x_n$在R^D空间内，分别具有类标签$c_1, c_2, \ldots, c_n$，NCA的目标是找到一种距离测度使近邻分类的效果尽可能最优。限定马氏距离变换矩阵Q是一个对称半正定矩阵，即$Q = A^T A$,那么可以得到两个样本点之间的马氏距离为：

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T Q (x_i - x_j)} = \sqrt{(Ax_i - Ax_j)^T (Ax_i - Ax_j)}$$

引入一个可微的softmax函数：

$$p_{ij} = \frac{e^{-\|Ax_i - Ax_j\|^2}}{\sum_{k \neq i} e^{-\|Ax_i - Ax_k\|^2}}, p_{ii} = 0$$

其中p_{ij}定义为：样本点x_i随机选择一个近邻，它最终选择另一个样本点x_j作为其近邻继而继承其类标签c_j的概率。这样，样本点x_i被正确分类的概率为：

$$p_i = \sum_{j \in C_i} p_{ij}, \ C_i = \{j | c_i = c_j\}$$

则目标函数要使得正确分类点的数据最大，定义为：

$$f(A) = \sum_i \sum_{j \in C_i} P_{ij} = \sum_i p_i$$

这是一个无约束优化问题，可以通过共轭梯度法或者随机梯度法求出A,当A是方阵时，经过NCA学习后的维数保持不变。当A是$d \times D$的非方阵时，可以将样本降维到R^d空间。

4.LDA、NCA同PCA的比较

(a) LDA和PCA的比较

相同点：LDA和PCA都是降维方法，两者的思想和计算方法非常类似，从几何的角度来看，PCA和LDA都是向低维空间做投影，都是将数据投影到新的相互正交的坐标轴上，最终的表现都是解一个矩阵特征值的问题。

不同点：**(1).**PCA是一种unsupervised的映射方法而LDA是一种supervised映射方法。LDA的输入数据是带标签的，而PCA的输入数据是不带标签的，也就是说PCA是一种无监督学习。**(2)**.PCA和LDA之间有着一些差异并适用于不同的应用场景。LDA通常来说是作为一个独立的算法存在，给定了训练数据后，将会得到一系列用于对新的输入产生判别的判别函数。而PCA更像是一个预处理的方法，它可以将原本的数据降低维度，而使得降低了维度的数据之间的方差最大。

(b) NCA和PCA的比较

相同点：NCA和PCA都是降维方法

不同点：**(1).**NCA算法本质是对样本进行线性或者非线性变换获取另一种更有类别区分度的表示形式，而PCA是将主成分维度组成的向量空间作为低维空间，将高维数据投影到这个空间。**(2).**使用NCA方法不仅可以进行降维也可以进行距离测度学习，没有很复杂的矩阵运算，也不需要对样本空间作分布作特定假设，而PCA只能用于降维，且涉及许多矩阵运算，一般而言也在假设数据分布满足高斯分布时才能得到最优的主元。

三、参考链接：

1.https://www.cnblogs.com/NextNight/p/6180542.html

2.https://www.cnblogs.com/LeftNotEasy/archive/2011/01/08/lda-and-pca-machine-learning.html

3.https://blog.csdn.net/chlele0105/article/details/13006443