1．LDA 与 PCA 的比较

PCA 降维是通过寻找特征方向， 将高维数据投影至特征方向， 在每个特征方向上将对应一个一维数据， 使用每个特征方向上对应的一维数据表示原始数据． 降维后的数据维度由特征方向个数决定，特征方向越多，降维后数据维度越高，降维后的数据与原始数据的差距越小；反之，特征方向个数越少，降维后数据维度越低，降维后的数据与原始数据的差距越大． 当特征方向的个数与原始数据维度相同时， PCA 降维后的数据与原始数据无差别，但是无法实现降维的作用．

LDA 同样是选择投影方向进行降维， 但它需要考虑样本的类别标记， 选择能将不同类数据分离开的投影方向；使得降维后同类数据尽可能集中，同时不同类数据尽可能分离．

PCA 与 LDA 的比较：

(1)相同点

·都是线性降维方法；

·都假设原始数据服从高斯分布；

·求解过程都应用特征值分解；

(2)不同点

·PCA 是无监督的降维方法，不考虑样本类别信息；LDA 是有监督的降维方法，需要考虑样本的类别信息；

·PCA 可以把 D 维数据降维至 0-D 维任意维度； LDA 只能将 D 维数据降维至(K-1)维，其中 K 是样本类别种类，无法降维至更低维度；

·PCA 只能用来进行数据降维， 它选择投影值方差尽可能大的方向进行降维，从而保留尽可能多的原始信息； LDA 在降维的同时， 还实现了数据分类的功能，LDA 降维的依据是样本类别信息，将不同类别的数据尽可能的分开；

2．NCA 与 PCA 的比较

NCA 是一种度量学习方法， 同时也可以进行数据降维． NCA 基于 KNN 实现， KNN 使用距离样本最近的 K 个邻居样本的类别信息决定该样本的类别．NCA 通过学习得到度量矩阵，使用度量矩阵计算两个样本间的距离， 代替 KNN 的欧氏距离； 同时根据每个样本与待分类样本的度量距离作为类别标记做参考类别的概率． 度量矩阵是 d*D 的矩阵， d 即降维后的数据维度．

PCA 与 NCA 的比较：

(1)相同点

·都可以进行数据降维

(2)不同点

·PCA 时无监督降维方法；NCA 是有监督的降维方法，同时还是度量学习方法；

·PCA 要求数据服从高斯分布；NCA 对数据分布没有要求；

·PCA 选择投影值方差尽可能大的方向进行降维，它的目的是保留尽可能多的原始信息；NCA 基于 KNN 分类方法选择最优的投影方向；

3．方法的原理

一、PCA 原理

PCA 是一种线性降维方法，它寻找原始数据的低维表示，并希望保留尽可能多的信息．降维方式为寻找一个投影方向， 将所有数据投影至此方向， 投影后的数据方差越大， 认为

保留的信息越多.

设数据集 $X=\{x_1,x_2,\dots,x_N\}$; $x_i\in \mathbb{R}^D$, $1\le i\le N$; $\mathbb{E}[X]=0$;

设投影方向为 w; x_i 在 w 方向的投影为

$$\frac{x^T w w}{\|w\|^2}.$$

由于 w 长度不影响分析，不妨设$\|w\|=1$，则x_i的投影值为

$$x^T w.$$

所有数据投影值的方差为

$$var(x^T w) = w^T cov(x) w,$$

其中$cov(x)=\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})^T$;

因此，寻找最优投影方向的问题转换为优化问题

$$\max_{w} \quad w^T cov(x) w \ \text{ s.t. } \|w\|=1.$$

使用拉格朗日乘子法可知，该优化问题最优解的必要条件是$cov(x)w=\lambda w$；将此条件带入方差项，可得

$$var(xTw)=\lambda$$

对协方差矩阵$cov(x)$进行谱分解，最大特征值对应得特征向量即为所求方向.

二、LDA 原理

LDA 也是线性降维方法， 它寻找有类别信息的高维数据的低维表示， 并且希望降维表示能体现类别信息. 以用于二分类的 FLD 为例，降维后同类数据越密集越好，不同类数据越分散越好.

设样本集合为$X=\{x_1,x_2,\dots,x_N\}$; $x_i\in\mathbb{R}^D, 1\le i\le N$; 其中,正样本集合为 X_1,$N_1=|X_1|$; 负样本集合为 X_2,$N_2=|X_2|$;

正样本均值，负样本均值分别为

$$\boldsymbol{m}_1=\frac{1}{N_1}\sum_{x\in X_1}x$$

$$\boldsymbol{m}_2=\frac{1}{N_2}\sum_{x\in X_2}x$$

正样本协方差矩阵，负样本协方差矩阵分别为

$$C_1=\frac{1}{N_1}\sum_{x\in X_1}(x-\boldsymbol{m}_1)(x-\boldsymbol{m}_1)^T$$

$$C_2=\frac{1}{N_2}\sum_{x\in X_2}(x-\boldsymbol{m}_2)(x-\boldsymbol{m}_2)^T$$

设 w 为投影方向，则

正样本投影均值，负样本投影均值分别为

$$m_1=\boldsymbol{m}_1^T w$$

$$m_2=\boldsymbol{m}_2^T w$$

正样本投影值方差，负样本投影值方差分别为

$$w^T C_1 w$$

$$w^T C_2 w$$

目标是使得m_1-m_2的绝对值尽可能大，同时$w^T(C_1+C_2)w$尽可能小;

将目标函数设置为

$$\frac{w^T(\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T w}{w^T(C_1+C_2)w}$$

定义类间散度矩阵和类内散度矩阵分别为

$$s_B=(\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T$$
$$S_w=N_1C_1+N_2C_2$$

则目标函数定义为

$$J=\frac{w^T s_B w}{w^T s_w w}$$

最优解的必要条件是

$$\frac{\partial J}{\partial w}=2\left(\frac{(w^T s w^w)s_B w-(w^T s_B w)s_w w}{(w^T s_w w)^2}\right)=0$$

即

$$S_w w=\frac{w^T s_w w}{w^T s_B w}(\boldsymbol{m}_1-\boldsymbol{m}_2)^T w(\boldsymbol{m}_1-\boldsymbol{m}_2)=c(\boldsymbol{m}_1-\boldsymbol{m}_2)$$

所以最优解的投影方向为

$$S_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)$$

在多分类 FLD 中，需要重新定义s_B,s_w，并求解广义特征值问题$s_B w=\lambda s_w w$

三、NCA 原理

NCA 是一种度量算法，基于以马氏距离为度量的 KNN 算法，以迭代方式学习转换矩阵.

设样本集合为$X=\{x_1,x_2,\dots,x_N\}$； $x_i\in\mathbb{R}^D,1\le i\le N$；

设转换矩阵为 A，A 是 d*D 的矩阵，d 即降维后的维度.

定义 softmax 函数为

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

p_{ij}代表样本x_i继承样本x_j类别的概率.

样本x_i能正确分类的概率为

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

其中C_i表示和样本x_i的真实类别有同样类别的样本集合.

定义目标函数为

$$f(A)=\sum_i p_i$$

梯度为

$$\frac{\partial f}{\partial A}=2A\sum_i\left(p_i\sum_k P_{ik}(x_i-x_k)(x_i-x_k)^T-\sum_{j\in C_i}P_{ij}\left(x_i-x_j\right)\left(x_i-x_j\right)^T\right)$$

使用梯度下降法迭代计算 A.