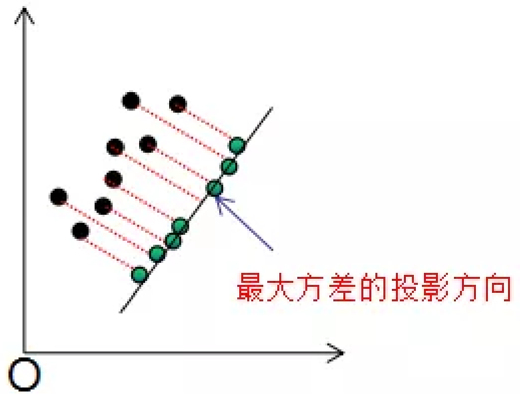
局部线性嵌入（LLE）原理总结

原创 石头 机器学习算法那些事 2019-03-22

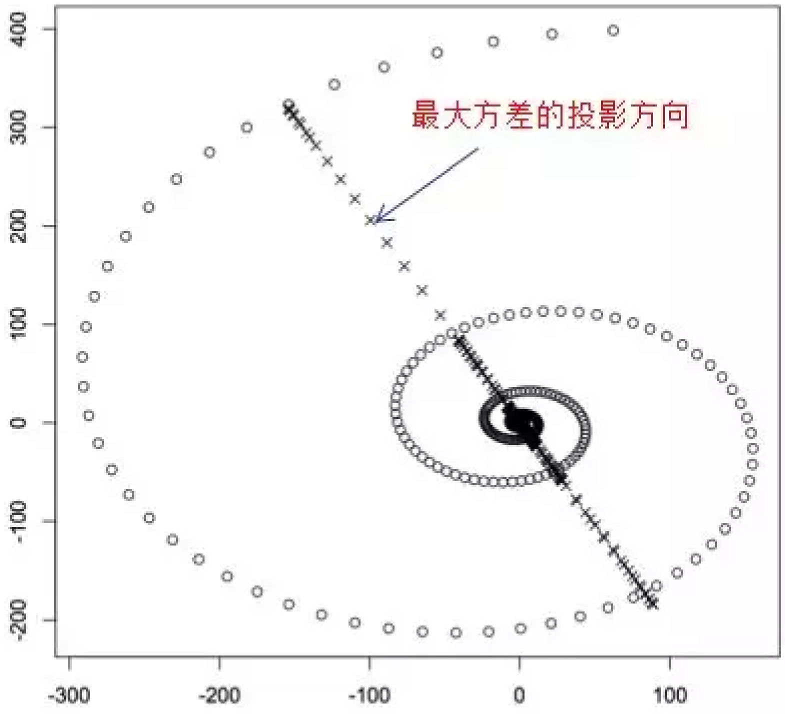
相关文章推荐：
主成分分析（PCA）原理总结
奇异值分解（SVD）原理总结
scikit-learn中PCA的使用方法
LLE参考资料下载：
链接:
https://pan.baidu.com/s/1Ht01oCGpUjL7rCX5bAhnYg
提取码: 62wq

LLE是一种非监督式的降维方法，与PCA降维方法相比，LLE降维后保持了初始样本间的局部关系。PCA是基于投影的最大方差的线性投影，在介绍LLE原理之前，我们首先用几张图看看PCA的降维效果：

对于线性数据，PCA很好的保留了初始样本的结构信息：

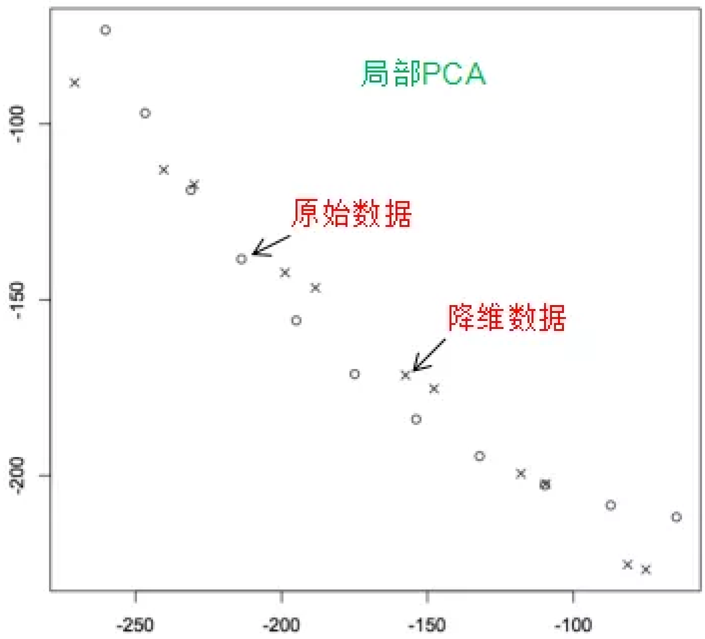


对于流型数据，PCA降维效果如下图：

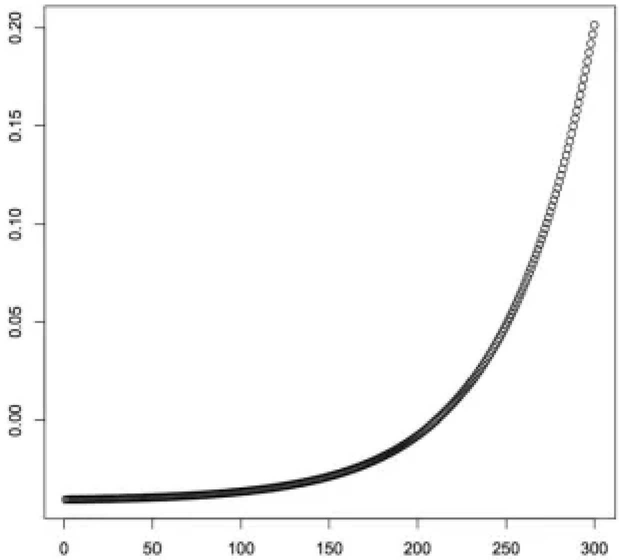


由上图可知，当初始样本集是一种螺旋形状的非线性结构，经过PCA降维后，结构已经发生了极大的改变，线性降维无法表达螺旋这种非线性结果，因此，PCA对于这类流行数据的降维效果很差。

若我们用积分思想去对这类流行数据进行PCA降维，对流形数据进行分段，每一段运用PCA降维，局部曲线曲度比较小，接近于直线，因此局部PCA可以很好地拟合曲线：



LLE的核心思想是依次对分段数据进行PCA降维，效果图如下：

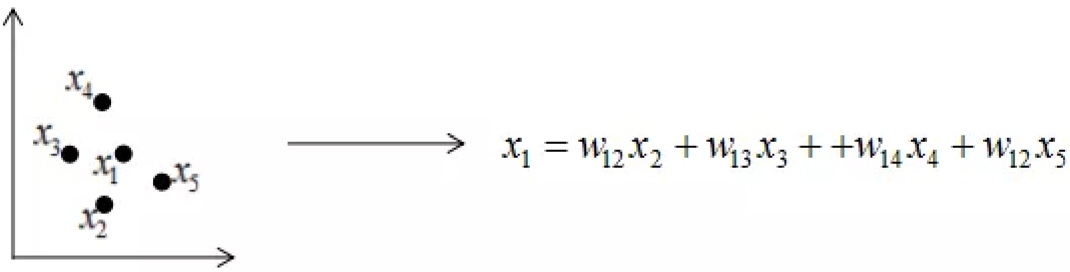


因此，LLE可以很好的拟合了原始样本集的结构信息 。

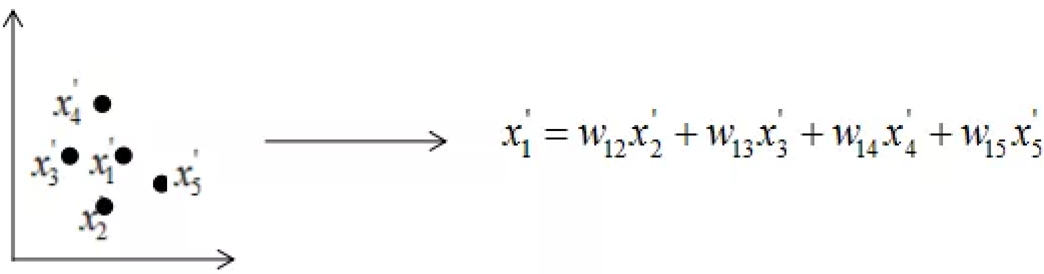
目录

1. LLE算法原理

LLE是Locally Linear Embedding的缩写，意思为局部线性嵌入。
局部线性的含义是整个数据集的局部范围内，数据是符合线性的，如下图：



嵌入的含义是降维，LLE的降维思想是不改变原始样本集的结构信息，即不改变对应的样本权重，如下图：



因此，投影前后的样本之间的关系（权重）不变，下面基于LLE思想去推导LLE算法过程：

2. LLE算法推导

推导算法前，我们首先假设样本和权重的矩阵表示：

初始样本集：$X_{D\times N}=[x_1,x_2,\dots,x_N]$

样本的K近邻权重：$W_{K\times N}=[w_1,w_2,\dots,w_N]$

LLE降维后的样本集：$Y_{d\times N}=[y_1,y_2,\dots,y_N]$

算法推导思想：

（1）根据初始样本集得到局部线性权重W：

$$RSS_i(w)=\sum_{i=1}^{N}\| x_i-\sum_{j=1}^{K}w_{ji}x_{ji}\|^2 \qquad (1)$$

最小化上式，可得每个样本的K近邻权重，W的定义：

$$W_{K\times N}=(w_1,w_2,\dots,w_N)$$

其中 $w_i=(w_{i1},w_{i2},\dots,w_{ik})^T$ 表示样本xi的K近邻权重的列向量，x_{ji} 表示xi的第j个近邻样本，w_{ji} 表示xi的第j个近邻样本的权重，对权重进行归一化：

$$\sum_{j=1}^{K}w_{ji}=1$$

（2）已知K近邻样本权重W，求解LLE降维的d维列向量 y_i：

$$\Phi(Y)=\sum_{i=1}^{N}\| y_i-\sum_{j=1}^{K}w_{ji}y_{ji}\|^2 \qquad (2)$$

最小化上式，可得降维后的样本数据集Y：

$$Y_{d\times N}=[y_1,y_2,\dots,y_N]$$

其中 y_i 表示样本降维后的列向量，y_{ji} 表示降维样本yi的第j个K近邻样本。我们对降维后的样本集进行标准化，即降维后的样本特征均值为0，标准差为1，特征间的协方差为0。用下式表示：

$$\sum_{i=1}^{N}y_i=0$$

$$\frac{1}{N}\sum_{i=1}^{N}y_iy_i^T=I_{d\times d}$$

好了，先停顿一下，看到这里你懂了这些数据集表示的含义了吗？

下面根据推导思想来求解LLE降维后样本集：

Step1：

我们首先通过KNN算法确定多少个邻域样本线性表示某个样本，假设这个值为K，则xi的K近邻点为：

$$N_i = KNN(x_i, K), N_i = [x_{1i}, x_{2i}, \dots, x_{Ki}] \qquad (3)$$

Step2：

最小化（1）式得到权重系数矩阵W，下面推导求解系数矩阵W的过程：

$$RSS(w_i) = \sum_{i=1}^{N} \| x_i - \sum_{j=1}^{K} w_{ji} x_{ji} \|^2 \qquad (1)$$

$$\because \quad \sum_{j=1}^{K} w_{ji} = 1$$

（1）式得：

$$\begin{aligned} &\Rightarrow \sum_{i=1}^{N} \| \sum_{j=1}^{K} w_{ji} x_i - \sum_{j=1}^{K} w_{ji} x_{ji} \|^2 \\ &= \sum_{i=1}^{N} \| \sum_{j=1}^{K} w_{ji} (x_i - x_{ji}) \|^2 \\ &= \sum_{i=1}^{N} \| (X_i - N_i) w_i \|^2 \qquad (4) \end{aligned}$$

其中$X_i = [\underbrace{x_i, x_i, \dots, x_i}_{k}]$，$N_i$的定义请参考式(3)

假设$\vec{r}$是列向量，那么$\| \vec{r} \|^2 = \vec{r}^T \vec{r}$

$\therefore$(4)式得：

$$\Rightarrow \sum_{i=1}^{N} w_i^T (X_i - N_i)^T (X_i - N_i) w_i$$

令$G_i = (X_i - N_i)^T (X_i - N_i)$，得：

$$\Rightarrow \sum_{i=1}^{N} w_i^T G_i w_i \qquad (5)$$

权重系数W满足约束条件：

$$\sum_{j=1}^{K} w_{ji} = 1$$

因此用拉格朗日乘子法最小化(5)式，得：

$$L(w_i,\lambda) = \sum_{i=1}^{N} w_i^T G_i w_i + \lambda(\mathbf{1}^T w_i)$$ ，其中$\mathbf{1}$表示$K\times 1$的全为1的列向量，$\mathbf{1} = [1,1,...1]^T_{K\times 1}$

令$\frac{\partial L}{\partial w_i} = 0$

得：

$$\frac{\partial L}{\partial w_i} = 2G_i w_i - \lambda \mathbf{1}^T = 0$$

$$\Rightarrow w_i = \frac{\lambda}{2} G_i^{-1}\mathbf{1}^T$$

λ的作用是使w_i的元素和为1

$$\therefore \quad w_i = \frac{G_i^{-1}\mathbf{1}}{\mathbf{1}^T G_i^{-1}\mathbf{1}}$$

即求得每个样本的权重系数W，

$$W = (w_1, w_2, ..., w_N)$$

因此求权重系数最关键的步骤是求Gi的逆矩阵

Step3：

已知权重系数，求LLE降维空间，即最小化下式：

$$\Phi(Y) = \sum_{i=1}^{N} \| y_i - \sum_{j=1}^{K} w_{ji} y_{ji} \|^2 \qquad (2)$$

约束条件：

$$\sum_{i=1}^{N} y_i = 0$$

$$\frac{1}{N}\sum_{i=1}^{N} y_i y_i^T = I_{d\times d}$$

对于不在K近邻范围内的样本，令该样本与中心样本的权重为零，因此，有：

$$w_{k+1,j} = w_{k+2,j} = w_{k+3,j} = ,... = w_{N,j} = 0$$

权重系数满足如下等式：

$$\sum_{j=1}^{K} w_{ji} = \sum_{j=1}^{N} w_{ji} = 1$$

令单位向量 $I_{N\times K} = [I_1, I_2, ..., I_K]$，权重矩阵 $W_{N\times K} = [\mathrm{w}_1, w_2, ..., w_k]$。

因此（2）式得：

$$\Phi(Y) = \sum_{i=1}^{N} \| Y(I_i - w_i) \|^2$$

矩阵 $A = [a_1, a_2, ..., a_N]$，有：

$$\sum_{\mathrm{i}=1}^{N} (a_i)^2 = \sum_{\mathrm{i}=1}^{N} a_i^T a_i = Tr(AA^T)$$

所以：

$$\Phi(Y) = tr(Y(I-W)(I-W)^T Y^T)$$

约束条件：$YY^T = \sum_{\mathrm{i}=1}^{N} y_i y_i^T = NI_{d\times d}$

令$M_{N\times N} = (I-W)(I-W)^T$，用拉格朗日乘子法求$Y$：

$$L(Y, \mathrm{u}) = YMY^T - u(YY^T - NI_{d\times d})$$

$$令\frac{\partial L(Y,u)}{\partial Y} = 2MY^T - 2uY^T = 0$$

$$\Rightarrow MY^T = uY^T$$

所以，Y的含义其实就是M的特征向量，LLE是从D维映射到d维，要求 $\Phi(Y)$的最小值，只要取M矩阵最小的d个特征值对应的特征向量即可，特征向量组成矩阵Y。

一般取最小的d个特征值时，不考虑最小的特征值。因此取特征值 $\mathrm{d}_2 \sim \mathrm{d}_{k+1}$ 对应的特征向量 $y_2 \sim y_{k+1}$，因此矩阵 $Y = [y_2, y_3, ..., y_{k+1}]$。

简单分下为什么不取最小特征值d1对应的特征向量？

$$W^T\vec{e}=\vec{e}$$

$$(W^T-I)\vec{e}=0$$

$$\because \vec{e}\text{是全为1的向量}$$

$$\therefore W^T-I=0$$

$$\text{两边转置}:(I-W)^T=0$$

$$\text{两边同时左乘}I-W:$$

$$(I-W)(I-W)^T=0$$

$$(I-W)(I-W)^T\vec{e}=0\vec{e}$$

所以矩阵M的最小特征值为0，一般不考虑特征值为0的特征向量，因为该特征值并不能反映数据特征。

3. LLE算法流程

1）确定每个样本的K近邻空间

2）求协方差矩阵：

$$G_i=(X_i-N_i)^T(X_i-N_i)$$

其中 $X_i=[\underbrace{x_i,x_i,\dots,x_i}_{k}]$，

$$N_i=[x_{1i},x_{2i},\dots,x_{Ki}]$$

3）根据协方差矩阵Gi，求解权重wi：

$$w_i=\frac{G_i^{-1}\vec{1}}{\vec{1}^TG_i^{-1}\vec{1}}$$

求每个样本的权重矩阵，得到W：

$$W=(w_1,w_2,\dots,w_N)$$

4）求矩阵M的最小d各特征值对应的特征向量yi，矩阵M：

$$M_{N\times N}=(I-W)(I-W)^T$$

其中矩阵$I_{N\times K}$是单位矩阵。

特征向量就是LLE输出的低维样本集矩阵Y：

$$Y=[y_2,y_3,\dots,y_{k+1}]$$

4. K近邻空间大于输入数据的维数D的情况分析

如果K>D，那么用K近邻空间的样本线性表示中心样本时，存在无数个解空间。用简单的例子来解释这一结论，如方程：

$$\begin{cases} x1 + x2 + x3 = 0 \\ x2 + x3 = 0 \end{cases}$$

有无穷个解，三元变量表示K近邻数，两个方程表示数据维数D，因此K>D时，有无穷个解。这种情况使得优化问题无规则（irregular）和不适定（ill-posed）。

解决方法是增加L2正则化，即最小化下式：

$$RSS(w_i) = \sum_{i=1}^{N} \| x_i - \sum_{j=1}^{K} w_{ji} x_{ji} \|^2 + \alpha \sum_{j=1}^{K} w_{ji}^2$$

其他的求解步骤如上面所写的一样。

5.小结

LLE对于流型数据有很好的降维效果，但是流型数据不能是闭合的，且K近邻数的选择对结果的影响很大。

参考

https://blog.csdn.net/scott198510/article/details/76099630

https://www.cnblogs.com/pinard/p/6266408.html