用scikit-learn进行LDA降维

在线性判别分析LDA原理总结中，我们对LDA降维的原理做了总结，这里我们就对scikit-learn中LDA的降维使用做一个总结。

1. 对scikit-learn中LDA类概述

在scikit-learn中， LDA类是sklearn.discriminant_analysis.LinearDiscriminantAnalysis。那既可以用于分类又可以用于降维。当然，应用场景最多的还是降维。和PCA类似，LDA降维基本也不用调参，只需要指定降维到的维数即可。

2. LinearDiscriminantAnalysis类概述

我们这里对LinearDiscriminantAnalysis类的参数做一个基本的总结。

1）**solver**：即求LDA超平面特征矩阵使用的方法。可以选择的方法有奇异值分解"svd"，最小二乘"lsqr"和特征分解"eigen"。一般来说特征数非常多的时候推荐使用svd，而特征数不多的时候推荐使用eigen。主要注意的是，如果使用svd，则不能指定正则化参数**shrinkage**进行正则化。默认值是svd

2）**shrinkage**：正则化参数，可以增强LDA分类的泛化能力。如果仅仅只是为了降维，则一般可以忽略这个参数。默认是None，即不进行正则化。可以选择"auto",让算法自己决定是否正则化。当然我们也可以选择不同的[0,1]之间的值进行交叉验证调参。注意shrinkage只在solver为最小二乘"lsqr"和特征分解"eigen"时有效。

3）**priors**：类别权重，可以在做分类模型时指定不同类别的权重，进而影响分类模型建立。降维时一般不需要关注这个参数。

4）**n_components**：即我们进行LDA降维时降到的维数。在降维时需要输入这个参数。注意只能为[1,类别数-1)范围之间的整数。如果我们不是用于降维，则这个值可以用默认的None。

从上面的描述可以看出，如果我们只是为了降维，则只需要输入n_components,注意这个值必须小于“类别数-1”。PCA没有这个限制。

3. LinearDiscriminantAnalysis降维实例

在LDA的原理篇我们讲到，PCA和LDA都可以用于降维。两者没有绝对的优劣之分，使用两者的原则实际取决于数据的分布。由于LDA可以利用类别信息，因此某些时候比完全无监督的PCA会更好。下面我们举一个LDA降维可能更优的例子。

完整代码参加我的github: https://github.com/ljpzzz/machinelearning/blob/master/classic-machine-learning/lda.ipynb

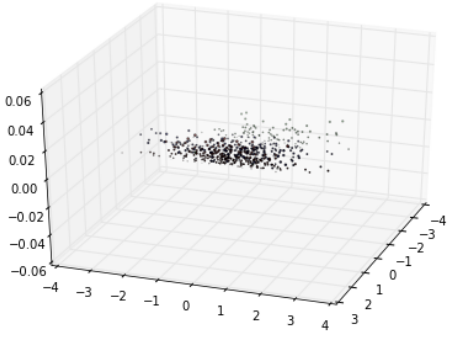
我们首先生成三类三维特征的数据，代码如下：

复制代码

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
%matplotlib inline
from sklearn.datasets.samples_generator import make_classification
X, y = make_classification(n_samples=1000, n_features=3, n_redundant=0, n_classes=3, n_informative=2,
                           n_clusters_per_class=1,class_sep =0.5, random_state =10)
fig = plt.figure()
ax = Axes3D(fig, rect=[0, 0, 1, 1], elev=30, azim=20)
plt.scatter(X[:, 0], X[:, 1], X[:, 2],marker='o',c=y)
```

复制代码

我们看看最初的三维数据的分布情况：



首先我们看看使用PCA降维到二维的情况，注意PCA无法使用类别信息来降维，代码如下：

复制代码

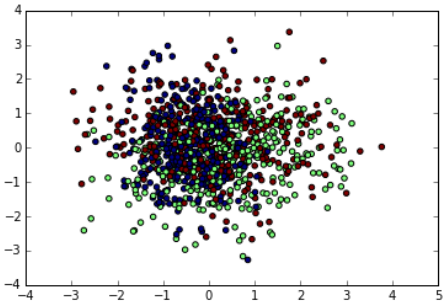
```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
pca.fit(X)
print pca.explained_variance_ratio_
print pca.explained_variance_
X_new = pca.transform(X)
plt.scatter(X_new[:, 0], X_new[:, 1],marker='o',c=y)
plt.show()
```

复制代码

在输出中，PCA找到的两个主成分方差比和方差如下：

```
[ 0.43377069  0.3716351 ]
[ 1.20962365  1.03635081]
```

输出的降维效果图如下:



由于PCA没有利用类别信息，我们可以看到降维后，样本特征和类别的信息关联几乎完全丢失。

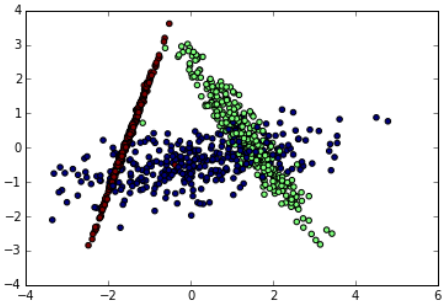
现在我们再看看使用LDA的效果，代码如下:

复制代码

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
lda = LinearDiscriminantAnalysis(n_components=2)
lda.fit(X,y)
X_new = lda.transform(X)
plt.scatter(X_new[:, 0], X_new[:, 1],marker='o',c=y)
plt.show()
```

复制代码

输出的效果图如下:



可以看出降维后样本特征和类别信息之间的关系得以保留。

一般来说，如果我们的数据是有类别标签的，那么优先选择LDA去尝试降维；当然也可以使用PCA做很小幅度的降维去消去噪声，然后再使用LDA降维。如果没有类别标签，那么肯定PCA是最先考虑的一个选择了。

（欢迎转载，转载请注明出处。欢迎沟通交流： liujianping-ok@163.com）

分类: 0081. 机器学习
标签: 维度规约
好文要顶 关注我 收藏该文
刘建平Pinard
关注 - 14
粉丝 - 2225
+加关注
4
0
« 上一篇：线性判别分析LDA原理总结
» 下一篇：奇异值分解(SVD)原理与在降维中的应用
posted @ 2017-01-04 17:04 刘建平Pinard 阅读(7783) 评论(15) 编辑 收藏

评论列表

#1楼 2017-01-04 18:58 codesnippet.info
文章一般，精神可嘉
支持(0)反对(1)

#2楼[楼主] 2017-01-05 10:11 刘建平Pinard
@ codesnippet.info
感谢回复，最近几个月在总结最近几年的机器学习相关的一些零碎的学习和项目的知识，欢迎批评指正。
支持(0)反对(0)

#3楼 2017-11-15 16:06 fns98
请问LDA的属性 coef_是什么值?
支持(0)反对(0)

#4楼[楼主] 2017-11-15 17:07 刘建平Pinard

@ fns98
你好，这个就是LDA投影矩阵WW。可以参看我写的原理篇：
http://www.cnblogs.com/pinard/p/6244265.html
支持(0)反对(0)

#5楼 2017-11-15 22:41 fns98
@ 刘建平Pinard
您好，我看了您的帖子，我还有些疑问
LDA可以进行降维也可以进行分类，那么作为分类的时候有是什么一个步骤呢？是否是可以这样理解：LDA作为分类器的时候，可以先将其降维，然后在低维空间假设数据满足高斯分布，然后得到高斯概率密度函数，根据类别概率分类。然后综上的所有叙述在sklearn中，就是一个fit和predict的两个函数。
刚刚接触LDA，对于总体框架有些疑惑，请大神赐教
支持(0)反对(0)

#6楼[楼主] 2017-11-16 11:34 刘建平Pinard
@ fns98
LDA分类的算法思路和降维的算法思路还是有很大区别的。在LDA分类算法中并没有先“将其降维，然后在低维空间假设数据满足高斯分布”。而是直接假设所有类别的数据一起符合多维高斯分布。
这块有机会我单独开一片细讲。
给你一个LDA分类的中文参考和英文参考：
http://blog.csdn.net/u014664226/article/details/52199892
http://web.stanford.edu/class/stats202/content/lec9.pdf
支持(1)反对(0)

#7楼 2017-11-16 14:04 fns98
@ 刘建平Pinard
谢谢！！
支持(0)反对(0)

#8楼 2017-12-19 17:27 lalalayujian
输入数据行列数 (663, 15)
LinearDiscriminantAnalysis(n_components=17, priors=None, shrinkage=None,
solver='svd', store_covariance=False, tol=0.0001)
降维后数据行列数 (663, 1)

你好，以上是我的输出结果。请问，我调用LDA降维（维度为17），输入的数据集只有15个变量，为什么不仅没有报错，而且降维后的数据怎么只有一列了啊？
支持(0)反对(0)

#9楼[楼主] 2017-12-20 10:29 刘建平Pinard
@ lalalayujian
你好，你的数据有多少种输出类别呢？注意最终降维到的维度数必须小于(类别数-1)。
如果你的训练数据的类别数小于19，那么你使用n_components=17，降维到17维是不行的。
支持(0)反对(0)

#10楼 2017-12-20 10:34 lalalayujian
@ 刘建平Pinard
嗯嗯，谢谢楼主了，原来是类别数-1，我以为是特征数-1。我的数据是两个类别标签，那输出结果就是对的。但是为什么参数里面n_components大于类别数-1了，它不报错呢
支持(0)反对(0)

#11楼[楼主] 2017-12-20 10:41 刘建平Pinard
@ lalalayujian
你好，sklearn的算法源程序估计没有对这一块做Validation。而LDA降维本身就是监督学习基于类别来做的，所以必须有这个限制，这个我在LDA的原理篇有讲到。

官方文档里面有提这个限制：
http://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html#sklearn.discriminant_analysis.LinearDiscriminantAnaly

n_components : int, optional
Number of components (< n_classes - 1) for dimensionality reduction.
支持(0)反对(0)

#12楼 2017-12-20 10:46 lalalayujian
嗯，谢谢楼主解惑了。我再研究一下原理
支持(0)反对(0)

#13楼 2018-04-28 16:02 涛涛不绝蕾蕾于冬
请问 shrinkage 这个参数真的是用于正则化吗?
http://scikit-learn.org/stable/modules/lda_qda.html 见上面链接的 1.2.4，怎么感觉讲的不是正则化?
支持(0)反对(0)

#14楼[楼主] 2018-04-30 23:28 刘建平Pinard
@ 涛涛不绝蕾蕾于冬
你好，这个参数在样本数比特征数还少的时候起作用，此时由于样本太少，正态样本分布不能很好的去拟合训练数据，也就是得到的协方差矩阵不能很好的反应训练数据特征，所以引入"shrinkage"来改善协方差矩阵的表达能力。其实也就是一种正则化的思想了，当然不像 L1，L2正则化那样明显罢了。
支持(0)反对(0)

#15楼 2018-08-01 15:04 小玲子zhl
@ codesnippet.info
站在说话不腰疼，你写过什么好的有价值的博客
支持(0)反对(0)

最新IT新闻:
· 怎样的物理学天才 让诺贝尔奖破例为他改了颁奖地点
· 硅谷是个什么谷（第1~4章）
· 马云放弃VIE所有权，是阿里治理的进步还是风险？
· 这家咖啡店能用个人信息换一杯免费咖啡，你愿意吗？
· 路易斯维尔大学学者解读诺奖：猛人PD-1的逆袭
» 更多新闻...

最新知识库文章:
· 为什么说 Java 程序员必须掌握 Spring Boot？
· 在学习中，有一个比掌握知识更重要的能力
· 如何招到一个靠谱的程序员
· 一个故事看懂“区块链”
· 被踢出去的用户
» 更多知识库文章...