



云服务器2核2G 低至49.68元/年

800GB月流量·随心用

立即抢购



园s.com

首页 新闻 博文 专区 闪存 班级

代码改变世界

刘建平Pinard

十五年码农，对数学统计学，数据挖掘，机器学习，大数据平台，大数据平台应用开发，大数据可视化感兴趣。

博客园 首页 新随笔 联系 订阅 管理

典型关联分析(CCA)原理总结

典型关联分析(Canonical Correlation Analysis，以下简称CCA)是最常用的挖掘数据关联关系的算法之一。比如我们拿到两组数据，第一组是人身高和体重的数据，第二组是对应的跑步能力和跳远能力的的数据。那么我们能不能说这两组数据是相关的呢？CCA可以帮助我们分析这个问题。

1. CCA概述

在数理统计里面，我们都知道相关系数这个概念。假设有两组一维的数据集X和Y，则相关系数 ρ 的定义为：

$$\rho(X,Y)=\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

其中 $cov(X,Y)$ 是X和Y的协方差，而 $D(X),D(Y)$ 分别是X和Y的方差。相关系数 ρ 的取值为[-1,1]， ρ 的绝对值越接近于1，则X和Y的线性相关性越高。越接近于0，则X和Y的线性相关性越低。

虽然相关系数可以很好的帮我们分析一维数据的相关性，但是对于高维数据就不能直接使用了。拿上面我们提到的，如果X是包括人身高和体重两个维度的数据，而Y是包括跑步能力和跳远能力两个维度的数据，就不能直接使用相关系数的方法。那我们能不能变通一下呢？CCA给了我们变通的方法。

CCA使用的方法是将多维的X和Y都用线性变换为1维的X'和Y'，然后再使用相关系数来看X'和Y'的相关性。将数据从多维变到1位，也可以理解为CCA是在进行降维，将高维数据降到1维，然后再用相关系数进行相关性的分析。下面我们看看CCA的算法思想。

2. CCA的算法思想

上面我们提到CCA是将高维的两组数据分别降维到1维，然后用相关系数分析相关性。但是有一个问题是，降维的标准是如何选择的呢？回想下主成分分析PCA，降维的原则是投影方差最大；再回想下线性判别分析LDA，降维的原则是同类的投影方差小，异类间的投影方差大。对于我们的CCA，它选择的投影标准是降维到1维后，两组数据的相关系数最大。

现在我们具体来讨论下CCA的算法思想。假设我们的数据集是X和Y，X为 $n_1 \times m$ 的样本矩阵。Y为 $n_2 \times m$ 的样本矩阵。其中m为样本个数，而 n_1, n_2 分别为X和Y的特征维度。

对于X矩阵，我们将其投影到1维，或者说进行线性表示，对应的投影向量或者说线性系数向量为 a ，对于Y矩阵，我们将其投影到1维，或者说进行线性表示，对应的投影向量或者说线性系数向量为 b ，这样X,Y投影后得到的一维向量分别为X',Y'。我们有

$$X'=a^TX,Y'=b^TY$$

我们CCA的优化目标是最大化 $\rho(X',Y')$ 得到对应的投影向量 a,b ，即

$$\underbrace{arg\max}_{a,b}\frac{cov(X',Y')}{\sqrt{D(X')}\sqrt{D(Y')}}$$

在投影前，我们一般会把原始数据进行标准化，得到均值为0而方差为1的数据X和Y。这样我们有：

$$cov(X',Y')=cov(a^TX,b^TY)=E(<a^TX,b^TY>)=E((a^TX)(b^TY)^T)=a^TE(XY^T)b$$

$$D(X')=D(a^TX)=a^TE(XX^T)a$$

$$D(Y')=D(b^TY)=b^TE(YY^T)b$$

由于我们的X，Y的均值均为0，则

$$D(X)=cov(X,X)=E(XX^T),D(Y)=cov(Y,Y)=E(YY^T)$$

$$cov(X,Y)=E(XY^T),cov(Y,X)=E(YX^T)$$

令 $S_{XY}=cov(X,Y)$ ，则优化目标可以转化为：

$$\underbrace{arg\max}_{a,b}\frac{a^TS_XYb}{\sqrt{a^TS_XXa}\sqrt{b^TS_YYb}}$$

由于分子分母增大相同的倍数，优化目标结果不变，我们可以采用和SVM类似的优化方法，固定分母，优化分子，具体的转化为：

$$\underbrace{arg\max}_{a,b}a^TS_XYb\\s.t.a^TS_XXa=1,b^TS_YYb=1$$

也就是说，我们的CCA算法的目标最终转化为一个凸优化过程，只要我们求出了这个优化目标的最大值，就是我们前面提到的多维X和Y的相关性度量，而对应的 a,b 则为降维时的投影向量，或者说线性系数。

公告

★珠江追梦，饮岭南茶，恋鄂北情
你的支持是我写作的动力：



昵称：刘建平Pinard
园龄：6年1个月
粉丝：9629
关注：15
+加关注

积分与排名

积分 - 487966
排名 - 1185

随笔分类 (135)

- 0040. 数学统计学(9)
- 0081. 机器学习(71)
- 0082. 深度学习(11)
- 0083. 自然语言处理(23)
- 0084. 强化学习(19)
- 0121. 大数据挖掘(1)
- 0122. 大数据平台(1)

随笔档案 (135)

- 2019年7月(1)
- 2019年6月(1)
- 2019年5月(2)
- 2019年4月(3)
- 2019年3月(2)
- 2019年2月(2)
- 2019年1月(2)
- 2018年12月(1)
- 2018年11月(1)
- 2018年10月(3)
- 2018年9月(3)
- 2018年8月(4)
- 2018年7月(3)
- 2018年6月(3)
- 2018年5月(3)
- 更多

常去的机器学习网站

强化学习入门书
52 NLP
Analytics Vidhya
深度学习进阶书
深度学习入门书
机器学习路线图
机器学习库

这个函数优化一般有两种方法，第一种是奇异值分解SVD，第二种是特征分解，两者得到的结果一样，下面我们分别讲解。

3. CCA算法的SVD求解

对于上面的优化目标，我们可以做一次矩阵标准化，就可以用SVD来求解了。

首先，我们令 $a = S_{XX}^{-1/2}u$, $b = S_{YY}^{-1/2}v$,这样我们有：

$$\begin{aligned}a^T S_{XX} a &= 1 \Rightarrow u^T S_{XX}^{-1/2} S_{XX} S_{XX}^{-1/2} u = 1 \Rightarrow u^T u = 1 \\ b^T S_{YY} b &= 1 \Rightarrow v^T S_{YY}^{-1/2} S_{YY} S_{YY}^{-1/2} v = 1 \Rightarrow v^T v = 1 \\ a^T S_{XY} b &= u^T S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2} v\end{aligned}$$

也就是说，我们的优化目标变成下式：

$$\begin{aligned}\underbrace{\arg \max}_{u,v} u^T S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2} v \\ s. t. u^T u = 1, v^T v = 1\end{aligned}$$

仔细一看，如果将u和v看做矩阵 $M = S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2}$ 的某一个奇异值对应的左右奇异向量。那么利用奇异值分解，我们可以得到 $M = U \Sigma V^T$,其中 U, V 分别为M的左奇异向量和右奇异向量组成的矩阵，而 Σ 为M的奇异值组成的对角矩阵。由于 U, V 所有的列都为标准正交基，则 $u^T U$ 和 $V^T v$ 得到一个只有一个标量值为1，其余标量值为0的向量。此时我们有

$$u^T S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2} v = u^T U \Sigma V^T v = \sigma_{uv}$$

也就是说我们最大化 $u^T S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2} v$,其实对应的最大值就是某一组左右奇异向量所对应的奇异值的最大值。也就是将M做了奇异值分解后，最大的奇异值就是我们优化目标的最大值，或者说我们的X和Y之间的最大相关系数。利用对应的左右奇异向量 u, v 我们也可以求出我们原始的X和Y的线性系数 $a = S_{XX}^{-1/2}u, b = S_{YY}^{-1/2}v$ 。

可以看出，SVD的求解方式非常简洁方便。但是如果你不熟悉SVD的话，我们也可以用传统的拉格朗日函数加上特征分解来完成这个函数的优化。

4. CCA算法的特征分解求解

特征分解方式就比较传统了，利用拉格朗日函数，优化目标转化为最大化下式：

$$J(a, b) = a^T S_{XY} b - \frac{\lambda}{2} (a^T S_{XX} a - 1) - \frac{\theta}{2} (b^T S_{YY} b - 1)$$

分别对 a, b 求导并令结果为0，我们得到：

$$\begin{aligned}S_{XY} b - \lambda S_{XX} a &= 0 \\ S_{YX} a - \theta S_{YY} b &= 0\end{aligned}$$

将上面第一个式子左乘 a^T ,第二个式子左乘 b^T ，并利用 $a^T S_{XX} a = 1, b^T S_{YY} b = 1$ ，我们得到

$$\lambda = \theta = a^T S_{XY} b$$

其实也就是说我们的拉格朗日系数就是要优化的目标。我们继续将上面的两个式子做整理，第一个式子左乘 S_{XX}^{-1} ,第二个式子左乘 S_{YY}^{-1} ，我们得到：

$$\begin{aligned}S_{XX}^{-1} S_{XY} b &= \lambda a \\ S_{YY}^{-1} S_{YX} a &= \lambda b\end{aligned}$$

将上面第二个式子带入第一个式子，我们得到

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX} a = \lambda^2 a$$

这个式子我们就熟悉了，这不就是特征分解吗！要求最大的相关系数 λ ,我们只需要对矩阵 $N = S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$ 做特征分解，找出最大的特征值取平方根即可，此时最大特征值对应的特征向量即为X的线性系数 a 。

同样的办法，我们将上面第一个式子带入第二个式子，我们得到

$$S_{YY}^{-1} S_{YX} S_{XX}^{-1} S_{XY} b = \lambda^2 b$$

，我们只需要对矩阵 $N' = S_{YY}^{-1} S_{YX} S_{XX}^{-1} S_{XY}$ 做特征分解，找出最大的特征值取平方根即可，此时最大特征值对应的特征向量即为Y的线性系数 b 。

可以看出特征分解的方法要比SVD复杂，但是两者求得的结果其实是等价的，只要利用SVD和特征分解之间的关系就很容易发现两者最后的结果相同。

5. CCA算法流程

这里我们对CCA的算法流程做一个总结，以SVD方法为准。

输入：各为m个的样本X和Y，X和Y的维度都大于1

输出：X,Y的相关系数 ρ ,X和Y的线性系数向量a和b

- 1) 计算X的方差 S_{XX} , Y的方差 S_{YY} , X和Y的协方差 S_{XY} , Y和X的协方差 $S_{YX} = S_{XY}^T$
- 2) 计算矩阵 $M = S_{XX}^{-1/2} S_{XY} S_{YY}^{-1/2}$

阅读

1. [线性判别分析LDA原理总结\(上\)](#)
2. [线性判别分析LDA原理总结\(下\)](#)
3. [线性判别分析LDA原理总结\(上\)](#)
4. [线性判别分析LDA原理总结\(下\)](#)
29. [线性判别分析LDA原理总结\(上\)](#)
5. [线性判别分析LDA原理总结\(下\)](#)

评论排行榜

1. [梯度提升树\(GBDT\)原理小结](#)
2. [集成学习之Adaboost算法原理](#)
3. [决策树算法原理\(下\)\(341\)](#)
4. [强化学习\(十六\) 深度确定性策略梯度](#)
5. [谱聚类 \(spectral clusterin](#)

推荐排行榜

1. [梯度下降 \(Gradient Descent\)](#)
2. [奇异值分解\(SVD\)原理与在降维中的应用](#)
3. [谱聚类 \(spectral clustering\)](#)
4. [集成学习之Adaboost算法原理](#)
5. [梯度提升树\(GBDT\)原理小结](#)

- 3) 对矩阵 M 进行奇异值分解, 得到最大的奇异值 ρ , 和最大奇异值对应的左右奇异向量 u, v
- 4) 计算X和Y的线性系数向量a和b, $a = S_{XX}^{-1/2}u, b = S_{YY}^{-1/2}v$

可见算法流程并不复杂, 但是要理解这个算法需要了解一些背景知识。

6. CCA算法小结

CCA算法广泛的应用于数据相关度的分析, 同时还是偏最小二乘法的基础。但是由于它依赖于数据的线性表示, 当我们的数据无法线性表示时, CCA就无法使用, 此时我们可以利用核函数的思想, 将数据映射到高维后, 再利用CCA的思想降维到1维, 求对应的相关系数和线性关系, 这个算法一般称为KCCA。

此外, 我们在算法里只找了相关度最大的奇异值或者特征值, 作为数据的相关系数, 实际上我们也可以像PCA一样找出第二大奇异值, 第三大奇异值, ...。得到第二相关系数和第三相关系数。然后对数据做进一步的相关性分析。但是一般的应用来说, 找出第一相关系数就可以了。

有时候我们的矩阵 S_{XX}, S_{YY} 不可逆, 此时我们得不到对应的逆矩阵, 一般遇到这种情况可以对 S_{XX}, S_{YY} 进行正则化, 将 S_{XX}, S_{YY} 变化为 $S_{XX} + \gamma I, S_{YY} + \gamma I$,然后继续求逆。其中 γ 为正则化系数。

(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: 关联算法

好文要顶

关注我

收藏该文







刘建平Pinard

粉丝 - 9629 关注 - 15

18

0

[+加关注](#)

« 上一篇: [用scikit-learn研究局部线性嵌入\(LLE\)](#)

» 下一篇: [Apriori算法原理总结](#)

posted @ 2017-01-16 17:19 刘建平Pinard 阅读(56475) 评论(59) 编辑 收藏 举报

[刷新评论](#) [刷新页面](#) [返回顶部](#)

登录后才能查看或发表评论, 立即 [登录](#) 或者 [逛逛](#) 博客园首页

- 【推荐】阿里云金秋云创季, 云服务器2核2G低至49.68元/年
- 【推荐】双十一同价! 腾讯云云服务器抢先购, 低至4.2元/月

编辑推荐:

· 为什么标准库的模板变量都是 inline 的

· .net 如何优雅的使用 EFCore

· 在 C# 中使用 Halcon 开发视觉检测程序

· 聊一聊如何截获 C# 程序产生的日志

· 一步一图带你深入理解 Linux 物理内存管理

阅读排行:

· .net如何优雅的使用EFCore

· 我开发的开源项目, 让.NET7中的EFCore更轻松地使用强类型Id

· 谁说.NET没有GC调优? 只改一行代码就让程序不再占用内存

· [Kogel.Subscribe.Mssql]SQL Server增量订阅, 数据库变更监听

· NET 6 实现滑动验证码 (三)、接口