奇異值分解(SVD)原理與在降維中的應用

奇異值分解(Singular Value Decomposition，以下簡稱SVD)是在機器學習領域廣泛應用的算法，它不光可以用於降維算法中的特徵分解，還可以用於推薦系統，以及自然語言處理等領域。是很多機器學習算法的基石。本文就對SVD的原理做一個總結，並討論在在PCA降維算法中是如何運用運用SVD的。

1. 回顧特徵值和特徵向量

我們首先回顧下特徵值和特徵向量的定義如下：

$$Ax = \lambda x$$

其中A是一個$n \times n$的實對稱矩陣，x是一個n維向量，則我們說λ是矩陣A的一個特徵值，而x是矩陣A的特徵值λ所對應的特徵向量。

求出特徵值和特徵向量有什麼好處呢？就是我們可以將矩陣A特徵分解。如果我們求出了矩陣A的n個特徵值$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$，以及這$n$個特徵值所對應的特徵向量$\{w_1, w_2, \ldots w_n\}$，如果這$n$個特徵向量線性無關，那麼矩陣A就可以用下式的特徵分解表示：

$$A = W\Sigma W^{-1}$$

其中W是這n個特徵向量所張成的$n \times n$維矩陣，而Σ為這n個特徵值為主對角線的$n \times n$維矩陣。

一般我們會把W的這n個特徵向量標準化，即滿足$||w_i||_2 = 1$，或者說$w_i^T w_i = 1$，此時W的n個特徵向量為標準正交基，滿足$W^TW = I$，即$W^T = W^{-1}$，也就是說W為酉矩陣。

這樣我們的特徵分解表達式可以寫成

$$A = W\Sigma W^T$$

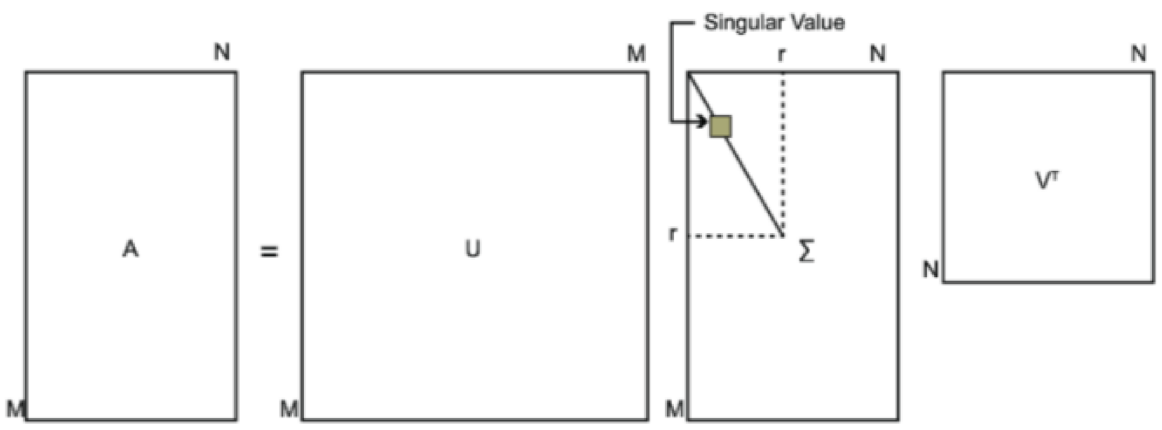
注意到要進行特徵分解，矩陣A必須為方陣。那麼如果A不是方陣，即行和列不相同時，我們還可以對矩陣進行分解嗎？答案是可以，此時我們的SVD登場了。

2. SVD的定義

SVD也是對矩陣進行分解，但是和特徵分解不同，SVD並不要求要分解的矩陣為方陣。假設我們的矩陣A是一個$m \times n$的矩陣，那麼我們定義矩陣A的SVD為：

$$A = U\Sigma V^T$$

其中U是一個$m \times m$的矩陣，Σ是一個$m \times n$的矩陣，除了主對角線上的元素以外全為0，主對角線上的每個元素都稱為奇異值，V是一個$n \times n$的矩陣。U和V都是酉矩陣，即滿足$U^TU = I, V^TV = I$。下圖可以很形象的看出上面SVD的定義：



那麼我們如何求出SVD分解後的U, Σ, V這三個矩陣呢？

如果我們將A的轉置和A做矩陣乘法，那麼會得到$n \times n$的一個方陣A^TA。既然A^TA是方陣，那麼我們就可以進行特徵分解，得到的特徵值和特徵向量滿足下式：

$$(A^TA)v_i = \lambda_i v_i$$

這樣我們就可以得到矩陣A^TA的n個特徵值和對應的n個特徵向量v了。將A^TA的所有特徵向量張成一個$n \times n$的矩陣V，就是我們SVD公式裡面的V矩陣了。一般我們將V中的每個特徵向量叫做A的右奇異向量。

如果我們將A和A的轉置做矩陣乘法，那麼會得到$m \times m$的一個方陣AA^T。既然AA^T是方陣，那麼我們就可以進行特徵分解，得到的特徵值和特徵向量滿足下式：

$$(AA^T)u_i = \lambda_i u_i$$

公告

★珠江追夢，飲嶺南茶，戀鄂北家★
你的支持是我寫作的動力：

暱稱： 劉建平Pinard
園齡： 3年5個月
粉絲： 5792
關注： 15
+加關注

積分與排名

積分- 461886
排名- 567

隨筆分類 (135)

0040. 數學統計學(9)
0081. 機器學習(71)
0082. 深度學習(11)
0083. 自然語言處理(23)
0084. 強化學習(19)
0121. 大數據挖掘(1)
0122. 大數據平台(1)

隨筆檔案 (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)
2017年8月(1)
2017年7月(3)
2017年6月(8)
2017年5月(7)
2017年4月(5)
2017年3月(10)
2017年2月(7)
2017年1月(13)
2016年12月(17)
2016年11月(22)

這樣我們就可以得到矩陣AA^T的m個特徵值和對應的m個特徵向量u了。將AA^T的所有特徵向量張成一個$m \times m$的矩陣U，就是我們SVD公式裡面的U矩陣了。一般我們將U中的每個特徵向量叫做A的左奇異向量。

U和V我們都求出來了，現在就剩下奇異值矩陣Σ沒有求出了。由於Σ除了對角線上是奇異值其他位置都是0，那我們只需要求出每個奇異值σ就可以了。

我們注意到：

$$A = U\Sigma V^T \Rightarrow AV = U\Sigma V^TV \Rightarrow AV = U\Sigma \Rightarrow Av_i = \sigma_i u_i \Rightarrow \sigma_i = Av_i / u_i$$

這樣我們可以求出我們的每個奇異值，進而求出奇異值矩陣Σ。

上面還有一個問題沒有講，就是我們說A^TA的特徵向量組成的就是我們SVD中的V矩陣，而AA^T的特徵向量組成的就是我們SVD中的U矩陣，這有什麼根據嗎？這個其實很容易證明，我們以V矩陣的證明為例。

$$A = U\Sigma V^T \Rightarrow A^T = V\Sigma^TU^T \Rightarrow A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^2V^T$$

上式證明使用了:$U^TU = I, \Sigma^T\Sigma = \Sigma^2$。可以看出$A^TA$的特徵向量組成的的確就是我們SVD中的V矩陣。類似的方法可以得到AA^T的特徵向量組成的就是我們SVD中的U矩陣。

進一步我們還可以看出我們的特徵值矩陣等於奇異值矩陣的平方，也就是說特徵值和奇異值滿足如下關係：

$$\sigma_i = \sqrt{\lambda_i}$$

這樣也就是說，我們可以不用$\sigma_i = Av_i / u_i$來計算奇異值，也可以通過求出A^TA的特徵值取平方根來求奇異值。

3. SVD計算舉例

這裡我們用一個簡單的例子來說明矩陣是如何進行奇異值分解的。我們的矩陣A定義為：

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$$

我們首先求出A^TA和AA^T

$$\mathbf{A^TA} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

$$\mathbf{AA^T} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

進而求出A^TA的特徵值和特徵向量：

$$\lambda_1 = 3; v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}; \lambda_2 = 1; v_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

接著求AA^T的特徵值和特徵向量：

$$\lambda_1 = 3; u_1 = \begin{pmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}; \lambda_2 = 1; u_2 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix}; \lambda_3 = 0; u_3 = \begin{pmatrix} 1/\sqrt{3} \\ -1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$$

利用$Av_i = \sigma_i u_i, i = 1, 2$求奇異值：

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \sigma_1\begin{pmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix} \Rightarrow \sigma_1 = \sqrt{3}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \sigma_2\begin{pmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{pmatrix} \Rightarrow \sigma_2 = 1$$

當然，我們也可以用$\sigma_i = \sqrt{\lambda_i}$直接求出奇異值為$\sqrt{3}$和1.

最終得到A的奇異值分解為：

$$A = U\Sigma V^T = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix}\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

4. SVD的一些性質

上面幾節我們對SVD的定義和計算做了詳細的描述，似乎看不出我們費這麼大的力氣做SVD有什麼好處。那麼SVD有什麼重要的性質值得我們注意呢？

2016年10月(8)

常去的機器學習網站

52 NLP
Analytics Vidhya
深度學習進階書
深度學習入門書
機器學習路線圖
機器學習庫
強化學習入門書

閱讀排行榜

1. 梯度下降（Gradient Descent）小結(276209)
2. 梯度提升樹(GBDT)原理小結(212244)
3. word2vec原理(一) CBOW與Skip-Gram模型基礎(176306)
4. 奇異值分解(SVD)原理與在降維中的應用(154915)
5. 線性判別分析LDA原理總結(144636)

評論排行榜

1. 梯度提升樹(GBDT)原理小結(492)
2. 集成學習之Adaboost算法原理小結(283)
3. 決策樹算法原理(下)(256)
4. word2vec原理(二) 基於Hierarchical Softmax的模型(250)
5. 譜聚類（spectral clustering）原理總結(221)

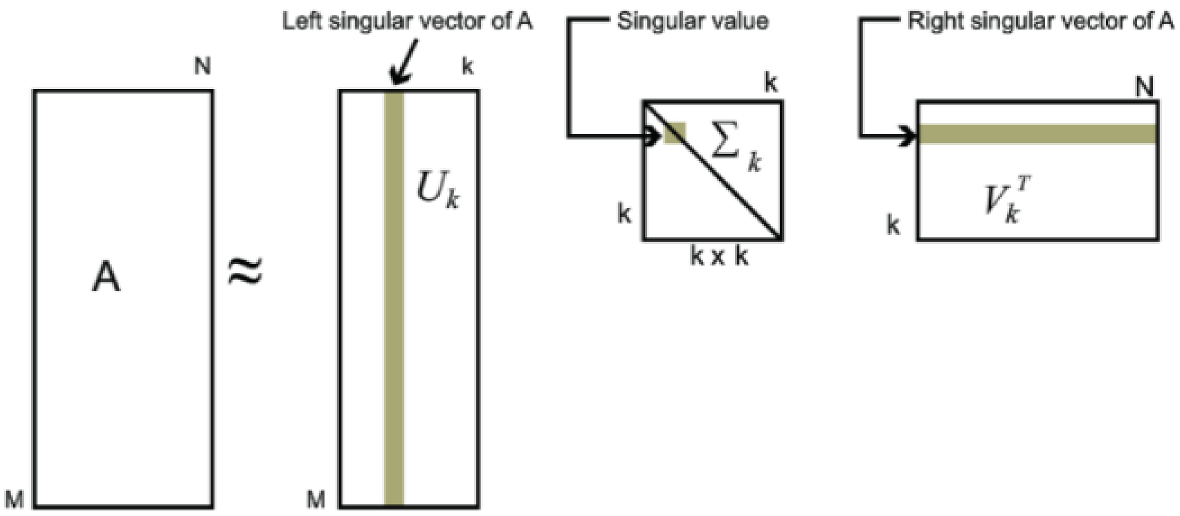
推薦排行榜

1. 梯度下降（Gradient Descent）小結(92)
2. 奇異值分解(SVD)原理與在降維中的應用(82)
3. 集成學習原理小結(40)
4. 梯度提升樹(GBDT)原理小結(38)
5. 譜聚類（spectral clustering）原理總結(37)

對於奇異值,它跟我們特徵分解中的特徵值類似，在奇異值矩陣中也是按照從大到小排列，而且奇異值的減少特別的快，在很多情況下，前10%甚至1%的奇異值的和就佔了全部的奇異值之和的99%以上的比例。也就是說，我們也可以用最大的k個的奇異值和對應的左右奇異向量來近似描述矩陣。也就是說：

$$A_{m \times n} = U_{m \times m}\Sigma_{m \times n}V^T_{n \times n} \approx U_{m \times k}\Sigma_{k \times k}V^T_{k \times n}$$

其中k要比n小很多，也就是一個大的矩陣A可以用三個小的矩陣$U_{m \times k}$, $\Sigma_{k \times k}$, $V^T_{k \times n}$來表示。如下圖所示，現在我們的矩陣A只需要灰色的部分的三個小矩陣就可以近似描述了。



由於這個重要的性質，SVD可以用於PCA降維，來做數據壓縮和去噪。也可以用於推薦算法，將用戶和喜好對應的矩陣做特徵分解，進而得到隱含的用戶需求來做推薦。同時也可以用於NLP中的算法，比如潛在語義索引（LSI）。下面我們就對SVD用於PCA降維做一個介紹。

5. SVD用於PCA

在主成分分析（PCA）原理總結中，我們講到要用PCA降維，需要找到樣本協方差矩陣X^TX的最大的d個特徵向量，然後用這最大的d個特徵向量張成的矩陣來做低維投影降維。可以看出，在這個過程中需要先求出協方差矩陣X^TX，當樣本數多樣本特徵數也多的時候，這個計算量是很大的。

注意到我們的SVD也可以得到協方差矩陣X^TX最大的d個特徵向量張成的矩陣，但是SVD有個好處，有一些SVD的實現算法可以不求先求出協方差矩陣X^TX，也能求出我們的右奇異矩陣V。也就是說，我們的PCA算法可以不用做特徵分解，而是做SVD來完成。這個方法在樣本量很大的時候很有效。實際上，scikit-learn的PCA算法的背後真正的實現就是用的SVD，而不是我們我們認為的暴力特徵分解。

另一方面，注意到PCA僅僅使用了我們SVD的右奇異矩陣，沒有使用左奇異矩陣，那麼左奇異矩陣有什麼用呢？

假設我們的樣本是$m \times n$的矩陣X，如果我們通過SVD找到了矩陣XX^T最大的d個特徵向量張成的$m \times d$維矩陣U，則我們如果進行如下處理：

$$X'_{d \times n} = U^T_{d \times m}X_{m \times n}$$

可以得到一個$d \times n$的矩陣X',這個矩陣和我們原來的$m \times n$維樣本矩陣X相比，行數從m減到了d，可見對行數進行了壓縮。也就是說，左奇異矩陣可以用於行數的壓縮。相對的，右奇異矩陣可以用於列數即特徵維度的壓縮，也就是我們的PCA降維。

6. SVD小結

SVD作為一個很基本的算法，在很多機器學習算法中都有它的身影，特別是在現在的大數據時代，由於SVD可以實現並行化，因此更是大展身手。SVD的原理不難，只要有基本的線性代數知識就可以理解，實現也很簡單因此值得仔細的研究。當然，SVD的缺點是分解出的矩陣解釋性往往不強，有點黑盒子的味道，不過這不影響它的使用。

（歡迎轉載，轉載請註明出處。歡迎溝通交流： liujianping-ok@163.com）

«上一篇： 用scikit-learn進行LDA降維
»下一篇： 局部線性嵌入(LLE)原理總結

< Prev 1 2 3

評論列表

#101樓 [樓主] 2019-10-27 11:02 劉建平Pinard

@ 李濤AT北京
你好，只能說經典的SVD用於PCA可能沒有優勢。

你說的求右奇異矩陣，需要求ATA的特徵向量，在很多SVD的實現算法庫是做了優化了，不需要按經典的思路來。

支持(0) 反对(0)

#102楼 [楼主] 2019-10-27 11:09 刘建平Pinard

@ lalalayujian
你好！
你理解的很对，按严格的数学定义来说，我这个平方写法是错误的。
这里的奇异值矩阵有个特性，比如你的$m > n$，那么最后的$m - n$行的值全部都是0，那么假如忽略这些捣乱的0，那么Σ就是一个方阵，就没有你说的问题了。

支持(0) 反对(0)

#103楼 2019-11-19 15:47 lalalayujian

@ 刘建平Pinard
你好，在计算出特征矩阵V后，我看有些资料计算Xnew = X *V^T(n,k)对吗，可是此处为何是V的转置取前k列呢，而不是Xnew = X *V(n,k)呢，V取前k列才是前k个特征向量呀？

支持(0) 反对(0)

#104楼 [楼主] 2019-11-20 09:01 刘建平Pinard

@ lalalayujian
你好，由于V是$n \times k$维度的，根据维度相容原理，那么$X_{new} = XV$即可。

如果某些文中定义的V维度是$k \times n$,这样才能加转置，但是这样的写法很少见。

支持(0) 反对(0)

#105楼 2019-11-20 20:29 zjdsk

博主你好，我有一个问题。假如A的秩为r，那ATA的秩应该也为r，那ATA最多只有r个特征值，也就是其特征向量只有r个，矩阵V是nxr的呀，怎么是nxn呢；矩阵U的维度同理。我思考了很久不知道自己哪里错了，希望博主能解答我的疑惑，谢谢

支持(0) 反对(0)

#106楼 [楼主] 2019-11-21 09:42 刘建平Pinard

@ zjdsk
你好！
看你是进行SVD还是进行SVD近似来降维了。

如果你只是进行SVD，那么无论你$A_{m \times n}$的秩是什么，最后V的维度都是$n \times n$,而不是$n \times r$

只有在你SVD近似来降维的时候，做了近似，那么此时V的维度才会是$n \times r$

支持(0) 反对(0)

#107楼 2019-12-16 10:43 dq116

屏幕不同缩放比会使公式错位，而且找不到一个使全部公式都不错位的缩放比...

支持(0) 反对(0)

#108楼 2020-03-07 18:36 才学疏浅的萝卜丝皮儿

@zjdsk还有0特征值对应的特征向量啊，前r个特征值向量对应的是非零特征值。

支持(0) 反对(0)

< Prev 1 2 3

刷新评论 刷新页面 返回顶部

注册用户登录后才能发表评论，请 登录 或 注册， 访问 网站首页。

相關博文：
· SVD（奇異值分解）Python實現

· 特徵值分解，奇異值分解（SVD）
· 特徵值分解與奇異值分解(SVD)
· SVD（奇異值分解）小結
· matlab特徵值分解和奇異值分解
» 更多推薦...

最新IT新聞：
· SpaceX載人龍飛船首次正式運營增加了NASA和日本JAXA宇航員
· Mojang新作《我的世界：地下城》5月28日發行
· 開放源代碼的項目Frontline Foods問世向醫院工作人員提供餐食
· Pokemon Go開發商Niantic收購3D世界掃描軟件公司6D.ai
· 研究人員首次對入侵癌細胞的物理力量進行了直接測量
» 更多新聞...