奇异值分解及几何意义 - 打不死的心态活到老 - CSDN博客

笔记本：	math
创建时间：	2017/9/24 13:01
作者：	87681654
URL：	http://blog.csdn.net/redline2005/article/details/24100293

奇异值分解及几何意义

分类：　matrix analysis and applicatio（2）

PS：一直以来对SVD分解似懂非懂，此文为译文，原文以细致的分析+大量的可视化图形演示了SVD的几何意义。能在有限的篇幅把这个如此清晰，实属不易。原文举了一个简单的图像处理问题，简单形象，真心希望路过的各路朋友能从不同的角度阐述下自己对SVD实际意比如 个性化推荐中应用了SVD，文本以及Web挖掘的时候也经常会用到SVD。

英文原文：We recommend a singular value decomposition

简介

SVD实际上是数学专业内容，但它现在已经渗入到不同的领域中。SVD的过程不是很好理解，因为它不够直观，但它对矩阵分解的效果却非常好。Netflix（一个提供在线电影租赁的公司）曾经就悬赏100万美金，如果谁能提高它的电影推荐系统评分预测准确率提高10%的话。令人惊讶的是，i了挑战，来自世界各地的团队运用了各种不同的技术。最终的获胜队伍"BellKor's Pragmatic Chaos"采用的核心算法就是基于SVD。

SVD提供了一种非常便捷的矩阵分解方式，能够发现数据中十分有意思的潜在模式。在这篇文章中，我们将会提供对SVD几何上的理解和一些简单

线性变换的几何意义(The geometry of linear transformations)

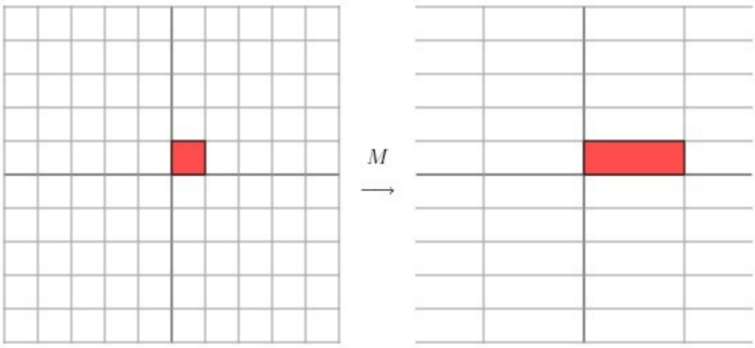
让我们来看一些简单的线性变换例子，以 2 X 2 的线性变换矩阵为例，首先来看一个较为特殊的，对角矩阵：

$$M = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

从几何上讲，M 是将二维平面上的点(x,y)经过线性变换到另外一个点的变换矩阵，如下图所示

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3x \\ y \end{bmatrix}.$$

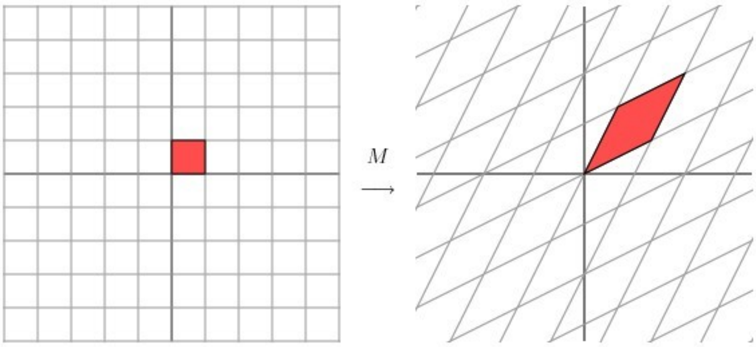
变换的效果如下图所示，变换后的平面仅仅是沿 X 水平方面进行了拉伸3倍，垂直方向是并没有发生变化。



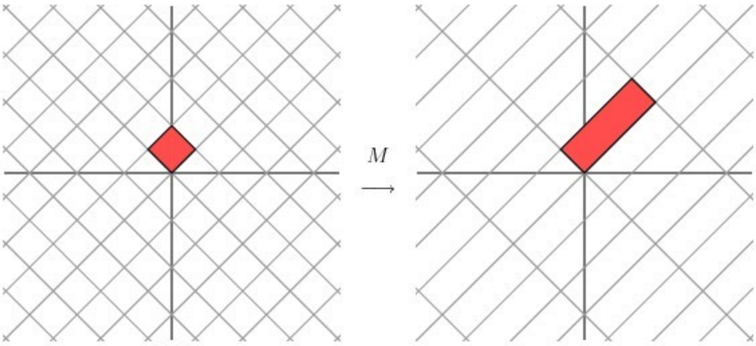
现在看下矩阵

$$M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

这个矩阵产生的变换效果如下图所示



这种变换效果看起来非常的奇怪，在实际环境下很难描述出来变换的规律 (这里应该是指无法清晰辨识出旋转的角度，拉伸的倍数之类的信息)。面的对称矩阵，假设我们把左边的平面旋转45度角，然后再进行矩阵M 的线性变换，效果如下图所示：



看起来是不是有点熟悉？ 对的，经过 **M** 线性变换后，跟前面的对角矩阵的功能是相同的，都是将网格沿着一个方向拉伸了3倍。

这里的 **M** 是一个特例，因为它是对称的。非特殊的就是我们在实际应用中经常遇见一些 非对称的，非方阵的矩阵。如上图所示，如果我们有一个矩阵**M** 的话，我们先将网格平面旋转一定的角度，**M** 的变换效果就是在两个维度上进行拉伸变换了。

用更加数学的方式进行表示的话，给定一个对称矩阵 **M** ，我们可以找到一些相互正交$\mathbf{v_i}$ ，满足$M\mathbf{v_i}$ 就是沿着$\mathbf{v_i}$ 方向的拉伸变换，公式如下：

$$M\mathbf{v_i} = \lambda_i\mathbf{v_i}$$

这里的 λ_i 是拉伸尺度(scalar)。从几何上看，**M** 对向量 $\mathbf{v_i}$ 进行了拉伸，映射变换。$\mathbf{v_i}$ 称作矩阵 M 的特征向量(eigenvector)，λ_i 称作为矩阵**M** 特征

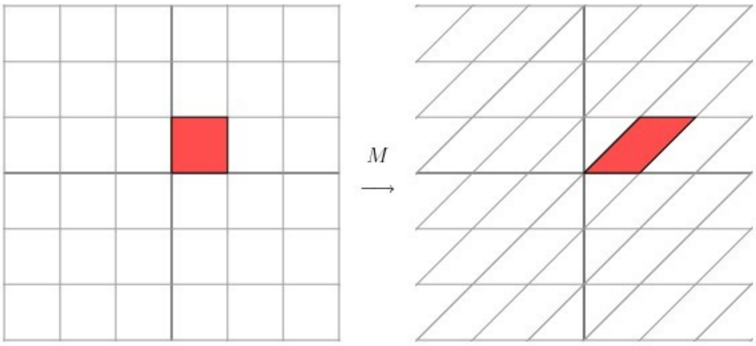
(eigenvalue)。这里有一个非常重要的定理，对称矩阵**M** 的特征向量是相互正交的。

如果我们用这些特征向量对网格平面进行线性变换的话，再通过 M 矩阵对网格平面进行线性换的效果跟对M 矩阵的特征向量进行线性变
一样的。

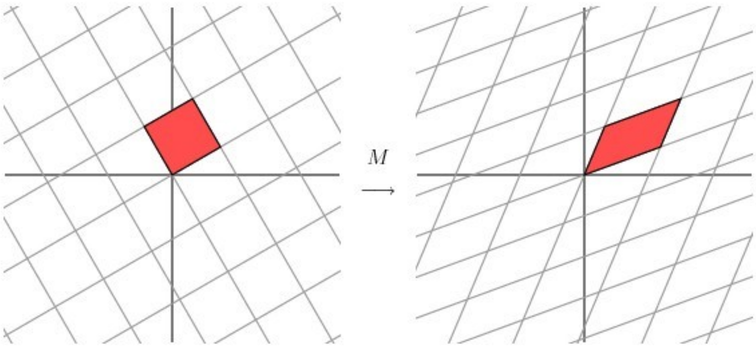
对于更为普通的矩阵而言，我们该怎么做才能让一个原来就是相互垂直的网格平面(orthogonal grid), 线性变换成另外一个网格平面同样垂直呢？**P**
直如图所示，就是两根交错的线条是垂直的。

$$M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

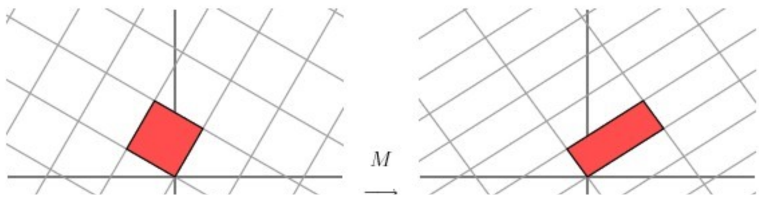
经过上述矩阵变换以后的效果如图

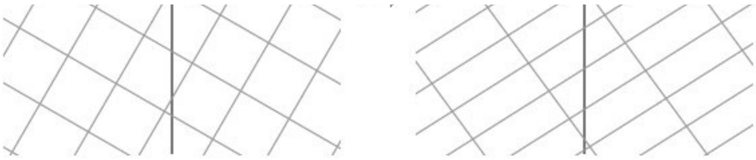


从图中可以看出，并没有达到我们想要的效果。我们把网格平面旋转 30 度角的话，然后再进行同样的线性变换以后的效果，如下图所示

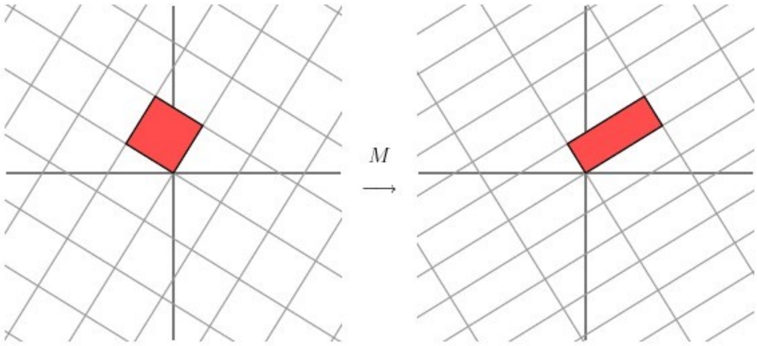


让我们来看下网格平面旋转60度角的时候的效果。





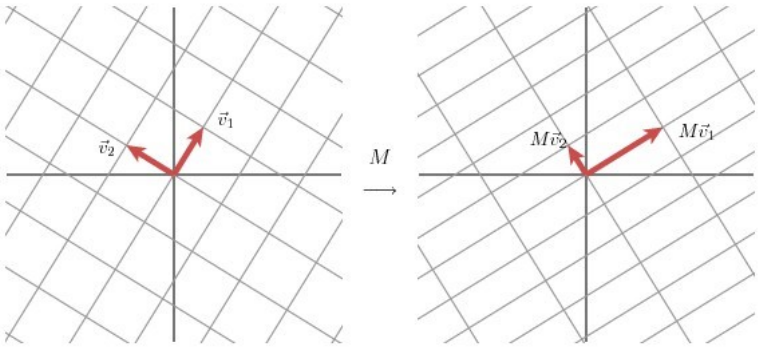
嗯嗯，这个看起来挺不错的样子。如果在精确一点的话，应该把网格平面旋转 58.28 度才能达到理想的效果。



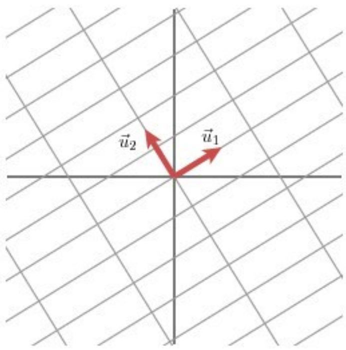
几何意义

该部分是从几何层面上去理解二维的SVD：对于任意的 2 x 2 矩阵，通过SVD可以将一个相互垂直的网格(orthogonal grid)变换到另外一个相互垂

我们可以通过向量的方式来描述这个事实: 首先，选择两个相互正交的单位向量 $\mathbf{v_1}$ 和 $\mathbf{v_2}$, 向量$M\mathbf{v_1}$ 和 $M\mathbf{v_2}$ 正交。



$\mathbf{u_1}$ 和$\mathbf{u_2}$分别表示$M\mathbf{v_1}$ 和 $M\mathbf{v_2}$的单位向量，$\sigma_1 * \mathbf{u_1} = M\mathbf{v_1}$ 和 $\sigma_2 * \mathbf{u_2} = M\mathbf{v_2}$。$\sigma_1$ 和 σ_2分别表示这不同方向向量上的模，也称作为矩阵M 的奇异



这样我们就有了如下关系式

$$M\mathbf{v_1} = \sigma_1\mathbf{u_1}$$

$$M\mathbf{v_2} = \sigma_2\mathbf{u_2}$$

我们现在可以简单描述下经过 M 线性变换后的向量 x 的表达形式。由于向量$\mathbf{v_1}$ 和$\mathbf{v_2}$是正交的单位向量，我们可以得到如下式子：

$$\mathbf{x} = (\mathbf{v_1} \cdot \mathbf{x})\mathbf{v_1} + (\mathbf{v_2} \cdot \mathbf{x})\mathbf{v_2}$$

这就意味着：

$$M\mathbf{x} = (\mathbf{v_1} \cdot \mathbf{x})M\mathbf{v_1} + (\mathbf{v_2} \cdot \mathbf{x})M\mathbf{v_2}$$

$$M\mathbf{x} = (\mathbf{v_1} \cdot \mathbf{x})\, \sigma_1\mathbf{u_1} + (\mathbf{v_2} \cdot \mathbf{x})\, \sigma_2\mathbf{u_2}$$

向量内积可以用向量的转置来表示，如下所示

$$\mathbf{v} \cdot \mathbf{x} = \mathbf{v}^T\mathbf{x}$$

最终的式子为

$$M\mathbf{x} = \mathbf{u_1}\sigma_1 \, \mathbf{v_1}^T\mathbf{x} + \mathbf{u_2}\sigma_2 \, \mathbf{v_2}^T\mathbf{x}$$

$$M = \mathbf{u_1}\sigma_1 \, \mathbf{v_1}^T + \mathbf{u_2}\sigma_2 \, \mathbf{v_2}^T$$

上述的式子经常表示成

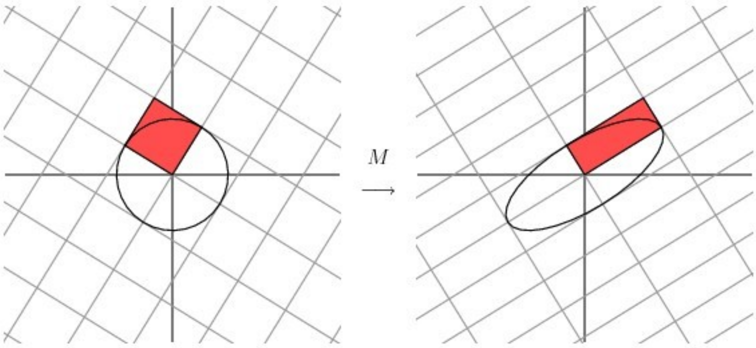
$$M = U\Sigma V^T$$

$\mathbf{u}$ 矩阵的列向量分别是$\mathbf{u_1}$,$\mathbf{u_2}$，Σ是一个对角矩阵，对角元素分别是对应的σ_1 和 σ_2，$\mathbf{V}$矩阵的列向量分别是$\mathbf{v_1}$,$\mathbf{v_2}$。上角标T 表示矩阵 $\mathbf{V}$ 的转置。

这就表明任意的矩阵 M 是可以分解成三个矩阵。$\mathbf{V}$表示了原始域的标准正交基，$\mathbf{u}$ 表示经过M 变换后的co-domain的标准正交基，Σ表示了$\mathbf{V}$ 中相对应向量之间的关系。(V describes an orthonormal basis in the domain, and U describes an orthonormal basis in the co-domain, and Σ desc much the vectors in V are stretched to give the vectors in U.)

如何获得奇异值分解？(How do we find the singular decomposition?)

事实上我们可以找到任何矩阵的奇异值分解，那么我们是如何做到的呢？假设在原始域中有一个单位圆，如下图所示。经过 M 矩阵变换以后在c 位圆会变成一个椭圆，它的长轴($M\mathbf{v_1}$)和短轴($M\mathbf{v_2}$)分别对应转换后的两个标准正交向量，也是在椭圆范围内最长和最短的两个向量。



换句话说，定义在单位圆上的函数$|M\mathbf{x}|$分别在$\mathbf{v_1}$和$\mathbf{v_2}$方向上取得最大和最小值。这样我们就把寻找矩阵的奇异值分解过程缩小到了优化函数$|M\mathbf{x}|$上（具体的推到过程这里就不详细介绍了）这个函数取得最优值的向量分别是矩阵 MT M 的特征向量。由于MTM是对称矩阵，因此不同特征值对应的是互相正交的，我们用vi 表示MTM的所有特征向量。奇异值$\sigma_i = |M\mathbf{v_i}|$， 向量 $\mathbf{u_i}$ 为 $M\mathbf{v_i}$ 方向上的单位向量。但为什么$\mathbf{u_i}$也是正交的呢？

推倒如下：

σ_i 和 σ_j分别是不同两个奇异值

$$M\mathbf{v_i} = \sigma_i\mathbf{u_i}$$

$$M\mathbf{v_j} = \sigma_j\mathbf{u_j}.$$

我们先看下$M\mathbf{v_i} \cdot M\mathbf{v_j}$，并假设它们分别对应的奇异值都不为零。一方面这个表达的值为0，推到如下

$$M\mathbf{v_i} \cdot M\mathbf{v_j} = \mathbf{v_i}^T M^T M\mathbf{v_j} = \mathbf{v_i} \cdot M^T M\mathbf{v_j} = \lambda_j \mathbf{v_i} \cdot \mathbf{v_j} = 0$$

另一方面，我们有

$$M\mathbf{v_i} \cdot M\mathbf{v_j} = \sigma_i\sigma_j \, \mathbf{u_i} \cdot \mathbf{u_j} = 0$$

因此，$\mathbf{u_i}$ 和$\mathbf{u_j}$是正交的。但实际上，这并非是求解奇异值的方法，效率会非常低。这里也主要不是讨论如何求解奇异值，为了演示方便，采用的都

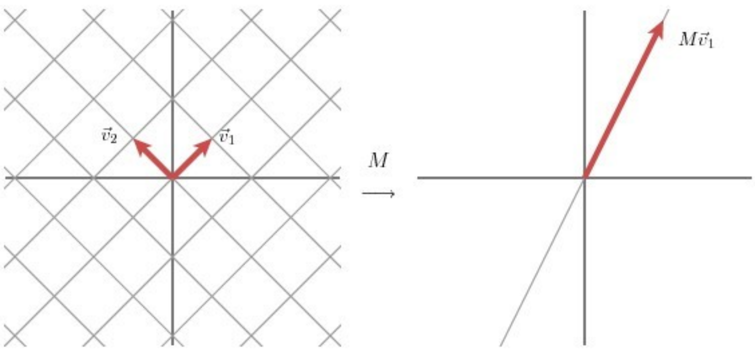
应用实例(Another example)

现在我们来看几个实例。

实例一

$$M = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

经过这个矩阵变换后的效果如下图所示



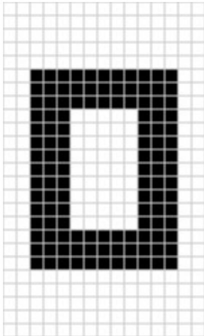
在这个例子中，第二个奇异值为 0，因此经过变换后只有一个方向上有表达。

$$M = \mathbf{u_1}\sigma_1 \mathbf{v_1}^T.$$

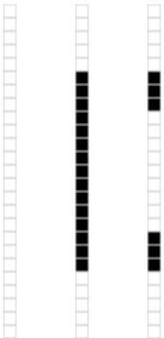
换句话说，如果某些奇异值非常小的话，其相对应的几项就可以不同出现在矩阵 M 的分解式中。因此，我们可以看到矩阵 M 的秩的大小等于非零数。

实例二

我们来看一个奇异值分解在数据表达上的应用。假设我们有如下的一张 15 x 25 的图像数据。



如图所示，该图像主要由下面三部分构成。



我们将图像表示成 15 x 25 的矩阵，矩阵的元素对应着图像的不同像素，如果像素是白色的话，就取 1，黑色的就取 0. 我们得到了一个具有375个如下图所示

$$M = \begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1
\end{bmatrix}$$

如果我们对矩阵M进行奇异值分解以后，得到奇异值分别是

$$\sigma_1 = 14.72$$

$$\sigma_2 = 5.22$$

$$\sigma_3 = 3.31$$

矩阵M就可以表示成

矩阵M就可以表示成

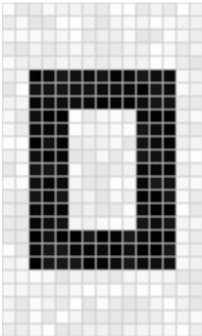
$$M=\mathbf{u}_1\sigma_1\,\mathbf{v}_1^T+\mathbf{u}_2\sigma_2\,\mathbf{v}_2^T+\mathbf{u}_3\sigma_3\,\mathbf{v}_3^T$$

$\mathbf{v}_i$具有15个元素，$\mathbf{u}_i$具有25个元素，σ_i 对应不同的奇异值。如上图所示，我们就可以用123个元素来表示具有375个元素的图像数据了。

实例三

减噪(noise reduction)

前面的例子的奇异值都不为零，或者都还算比较大，下面我们来探索一下拥有零或者非常小的奇异值的情况。通常来讲，大的奇异值对应的部分会息。比如，我们有一张扫描的，带有噪声的图像，如下图所示



我们采用跟实例二相同的处理方式处理该扫描图像。得到图像矩阵的奇异值：

$$\sigma_1 = 14.15$$

$$\sigma_2 = 4.67$$

$$\sigma_3 = 3.00$$

$$\sigma_4 = 0.21$$

$$\sigma_5 = 0.19$$

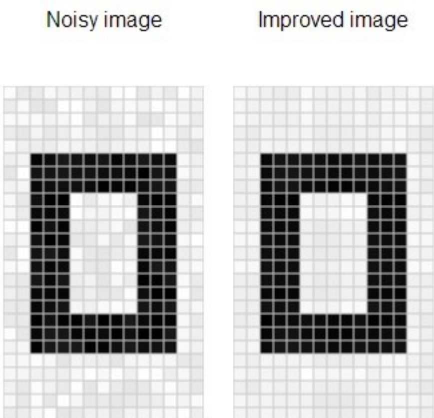
$$\ldots$$

$$\sigma_{15} = 0.05$$

很明显，前面三个奇异值远远比后面的奇异值要大，这样矩阵 M 的分解方式就可以如下：

$$M\approx\ \mathbf{u}_1\sigma_1\,\mathbf{v}_1^T+\mathbf{u}_2\sigma_2\,\mathbf{v}_2^T+\mathbf{u}_3\sigma_3\,\mathbf{v}_3^T$$

经过奇异值分解后，我们得到了一张降噪后的图像。

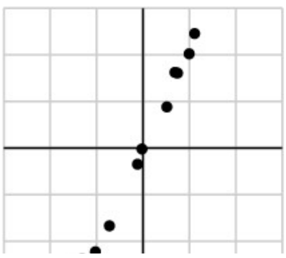


实例四

数据分析(data analysis)

我们搜集的数据中总是存在噪声：无论采用的设备多精密，方法有多好，总是会存在一些误差的。如果你们还记得上文提到的，大的奇异值对应了信息的话，运用SVD进行数据分析，提取其中的主要部分的话，还是相当合理的。

作为例子，假如我们搜集的数据如下所示：



我们将数据用矩阵的形式表示：

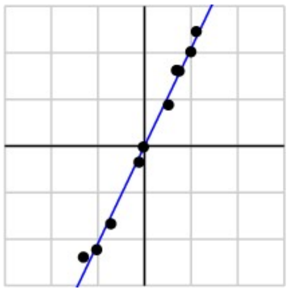
$$\begin{matrix} -1.03 & 0.74 & -0.02 & 0.51 & -1.31 & 0.99 & 0.69 & -0.12 & -0.72 & 1.11 \\ -2.23 & 1.61 & -0.02 & 0.88 & -2.39 & 2.02 & 1.62 & -0.35 & -1.67 & 2.46 \end{matrix}$$

经过奇异值分解后，得到

$$\sigma_1 = 6.04$$

$$\sigma_2 = 0.22$$

由于第一个奇异值远比第二个要大，数据中有包含一些噪声，第二个奇异值在原始矩阵分解相对应的部分可以忽略。经过SVD分解后，保留了主要示



就保留主要样本数据来看，该过程跟PCA(principal component analysis)技术有一些联系，PCA也使用了SVD去检测数据间依赖和冗余信息.

总结(Summary)

这篇文章非常的清晰的讲解了SVD的几何意义，不仅从数学的角度，还联系了几个应用实例形象的论述了SVD是如何发现数据中主要信息的。在许多团队都运用了矩阵分解的技术，该技术就来源于SVD的分解思想，矩阵分解算是SVD的变形，但思想还是一致的。之前算是能够运用矩阵分解推荐系统中，但理解起来不够直观，阅读原文后醍醐灌顶，我想就从SVD能够发现数据中的主要信息的思路，就几个方面去思考下如何利用数据中关系去探索个性化推荐系统。也希望路过的各位大侠不吝分享呀。

References:

Gilbert Strang,**Linear Algebra and Its Applications**. Brooks Cole

William H. Press*et al*,**Numercial Recipes in C: The Art of Scientific Computing**. Cambridge University Press.

Dan Kalman, A Singularly Valuable Decomposition: The SVD of a Matrix, *The College Mathematics Journal* **27** (1996), 2-23.

If You Liked This, You're Sure to Love That,*The New York Times*, November 21, 2008.

上一篇 A sigular value decomposition（奇异值分解）

下一篇 聚类算法K-Means, K-Medoids, GMM, Spectral clustering，Ncut .

相关文章推荐

- 奇异值分解(SVD)原理详解
- 奇异值分解SVD的理解与应用
- 携程机票大数据基础平台架构演进-- 许鹏
- Android Material Design 新控件
- 奇异值分解(SVD)原理详解及推导
- 矩阵特征值分解与奇异值分解含义解析及应用
- Python可以这样学--董付国
- 机器学习需要用到的数学知识
- 矩阵的特征值分解与奇异值分解的几何意义
- Singular Value Thresholding (SVT) 奇异值阈值
- 一步一步学Spring Boot
- 特征值分解、奇异值分解、PCA概念整理
- 机器学习 之 Haar特征
- SVD奇异值分解分析
- 深入浅出C++程序设计
- 深度学习与自然语言处理(1)_斯坦福cs224d Lect...