SVD Review

Mingdong Wu

Intro

SVD & PCA？ Singular vec？

降维？ 图像压缩？ Gx讲过SVD？

主成分是啥？

So on...

GX's SVD

设 A 是 $m \times n$ 实矩阵．由于 AA^T 实对称，

存在正交矩阵 P，使得

A 列向量的主方向，主...

$$AA^T = \underbrace{[\beta_1\ \beta_2 \cdots \beta_m]}_{P} \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & 0 \end{bmatrix} P^T$$

且 AA^T 的特征值非负：

$$\lambda_1 \geq \ldots \geq \lambda_r > \lambda_{r+1} = \ldots = 0$$

由

$$(A^T P)^T A^T P = P^T A A^T P = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_r & \\ & & & 0 \end{bmatrix}$$

知 $A^T P$ 的 m 个列向量 $A^T\beta_1, \ldots, A^T\beta_m$

两两正交且长度分别为

$$\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_r}, 0, \cdots, 0$$

GX's SVD

令 $\gamma_i = \dfrac{1}{\sqrt{\lambda_i}} A^T \beta_i$ ， $i = 1, \dots, r$;

则 $\gamma_1, \dots, \gamma_r$ 可扩充成 R^n 的标准正交基

$$\gamma_1, \dots, \gamma_r, \dots, \gamma_n$$

记正交矩阵 $Q = [\gamma_1 \ \dots \ \gamma_n]$，则有

$$A^T P = A^T [\beta_1 \dots \beta_m] = Q \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & \sqrt{\lambda_r} & \\ & & & 0 \end{bmatrix}_{n\times m}$$

即 $A = P S Q^T$

GX's SVD

Singular Value Decomposition

每个实矩阵 A 都能写成 $A = P\,S\,Q^T$

$$= [\,\beta_1 \;\; \beta_2 \cdots \beta_m\,] \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & \sqrt{\lambda_r} & \\ & & & 0 \end{bmatrix}_{m\times n} \begin{bmatrix} \gamma_1^T \\ \gamma_2^T \\ \vdots \\ \gamma_n^T \end{bmatrix}$$

$$= \sqrt{\lambda_1}\,\beta_1\,\gamma_1^T + \sqrt{\lambda_2}\,\beta_2\,\gamma_2^T + \cdots + \sqrt{\lambda_r}\,\beta_r\,\gamma_r^T$$

Two Perspective

- 下面我们从两种不同的视角分析SVD
- 一种是John的，将矩阵A视为向量簇
- 另一种是将矩阵A看成一个线性映射
- 咱先从John的开始

Task Definition

- 给定n个d维向量，求最优k($\leq r$)维子空间？
  - Linear
- 什么是“最优k维子空间”？
  - n个向量在该子空间上“**成分**”**最大**
- **什么是“成分”？**
  - **投影的Frobenius norm**
- Frobenius norm？

$$||A||_F = \sqrt{\sum_{j,k} a_{jk}^2}.$$

Task Definition

- 给定一个k-subspace，设它的一组标准正交基为$\{w_1, w_2 \dots w_k\}$
- 对向量簇$\{\alpha_1, \alpha_2 \dots \alpha_n\}$，令$A = [\alpha_1^T, \alpha_2^T \dots \alpha_n^T]^T$
- 则$\{\alpha_1, \alpha_2 \dots \alpha_n\}$到$\{w_1, w_2 \dots w_k\}$投影的Frobenius norm：

$$|A\mathbf{w_1}|^2 + |A\mathbf{w_2}|^2 + \cdots + |A\mathbf{w_{k-1}}|^2 + |A\mathbf{w_k}|^2$$

- 问题就转化为，给定A，找一组标准正交基$\{w_1, w_2 \dots w_k\}$，使上面的式子最大

Greedy Approach

- 如何找最优的$\{w_1, w_2 \ldots w_k\}$呢？先考虑一个朴素的想法：
- Step 1.先找一个v_1 ($|v_1|$=1)使$\{\alpha_1, \alpha_2 \ldots \alpha_n\}$到$v_1$的“成分”最大
  - John定义的“**成分**”**(Component)？就是“投影的” Frobenius norm**

$$\mathbf{v_1} = \arg\max_{|\mathbf{v}|=1} |A\mathbf{v}|. \qquad \sigma_1(A) = |A\mathbf{v_1}|$$

- 而找到的v_1就叫“(right)singular vector”，刻画的是“主方向”
- $\sigma_1^2(A)$就叫“singular value”，刻画的是向量簇在v_1上的“主成分”

Greedy Approach

- Step 2.再找v_2,使v_2和v_1正交的同时，$\{\alpha_1,\alpha_2 \ldots \alpha_n\}$到$v_1$的“成分”最大
- …
- Step m. 找v_{m},使v_{m}与$\{v_1,v_2 \ldots v_{\mathrm{m}-1}\}$正交同时，$\{\alpha_1,\alpha_2 \ldots \alpha_n\}$到$v_{\mathrm{m}}$的“成分”最大
- 当算法执行m > r步之后，会发现：$\max\limits_{\substack{\mathbf{v}\perp\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r \\ |\mathbf{v}|=1}} |A\mathbf{v}| = 0.$
- 算法自然停止

Greedy Approach Works

- 贪心找出来的k个奇异向量$\{v_1, v_2 \ldots v_k\}$为何恰好就是最优k-subspace？
- 归纳法证明：
  - K=1显然
  - K=k-1 时，假设命题成立，考虑K=k的情形
  - 如果$\{w_1, w_2 \ldots w_k\}$是最优的k-subspace的标准正交基
  - $\{v_1, v_2 \ldots v_{k-1}\}$是最优的(k-1)-subspace
  - 总能调整w_k，使其与空间$\{v_1, v_2 \ldots v_{k-1}\}$正交

$$|A\mathbf{w_k}|^2 \leq |A\mathbf{v_k}|^2.$$

$$|A\mathbf{w_1}|^2 + |A\mathbf{w_2}|^2 + \cdots + |A\mathbf{w_{k-1}}|^2 \leq |A\mathbf{v_1}|^2 + |A\mathbf{v_2}|^2 + \cdots + |A\mathbf{v_{k-1}}|^2$$

Singular Vectors and Values

- 回过头研究下我们找到的$\{v_1, v_2 \dots v_r\}$和$\{\sigma_1, \sigma_2 \dots \sigma_r\}$的性质
- 性质1. $\{\alpha_1, \alpha_2 \dots \alpha_n\}$可以被$\{v_1, v_2 \dots v_r\}$线性表出
- $\alpha_i = \sum_j (\alpha_i, v_j) v_j$
- 性质2. “A的成分” 也应等于 “A被分解后的成分”

$$\sum_{j=1}^{n} |\mathbf{a_j}|^2 = \sum_{j=1}^{n} \sum_{i=1}^{r} (\mathbf{a_j} \cdot \mathbf{v_i})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n} (\mathbf{a_j} \cdot \mathbf{v_i})^2 = \sum_{i=1}^{r} |A\mathbf{v_i}|^2 = \sum_{i=1}^{r} \sigma_i^2(A).$$

- 定义left singular vec:

$$\mathbf{u_i} = \frac{1}{\sigma_i(A)} A\mathbf{v_i}.$$

Singular Value Decomposition

- 给定矩阵A的right singular vec $\{v_1, v_2 \ldots v_{\mathrm{r}}\}$和left singular vec $\{\mathrm{u}_1, u_2 \ldots u_r\}$以及singular value $\{\sigma_1, \sigma_2 \ldots \sigma_r\}$

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T .$$

Singular Value Decomposition

- 证明John版的SVD $A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T.$
- Lemma：如果对任意v, Av = Bv, 那么A = B
- 对任意的v，如果它在$\{v_1, v_2 \dots v_r\}$张成的空间里：

$$\sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \mathbf{v_j} = \sigma_j \mathbf{u_j} = A\mathbf{v_j} \qquad A\mathbf{v} = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T \mathbf{v}$$

- 如果它和$\{v_1, v_2 \dots v_r\}$正交，那么 两端都为0(左右依旧相等)
- 利用lemma SVD得证
- 想想这和GX给出SVD的方式有什么不同？

Singular Value Decomposition

- 我知道怎么用John的算法求SVD，但有没有看起来更方便的算法？
- 我们构造个矩阵B

$$B = A^T A = \left(\sum_i \sigma_i \mathbf{v_i} \mathbf{u_i}^T\right)\left(\sum_j \sigma_j \mathbf{u_j} \mathbf{v_j}^T\right)$$

$$= \sum_{i,j} \sigma_i \sigma_j \mathbf{v_i} (\mathbf{u}_i^T \cdot \mathbf{u_j}) \mathbf{v_j}^T = \sum_i \sigma_i^2 \mathbf{v_i} \mathbf{v_i}^T.$$

- 可以发现B的特征值和特征向量恰好就是(right)singular vec 和 singular value：

$$B\mathbf{v_j} = \left(\sum_i \sigma_i^2 \mathbf{v_i} \mathbf{v_i}^T\right) \mathbf{v_j} = \sigma_j^2 \mathbf{v_j},$$

Truncation

- 有了SVD之后，我们能很自然地开始考虑k-部分和

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u_i} \mathbf{v_i}^T. \longrightarrow A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

- 性质1: A_k的行向量，正是A的行向量投影到$\{v_1, v_2 \ldots v_k\}$的结果

$$\mathbf{a} \longrightarrow \sum_{i=1}^{k} (\mathbf{a} \cdot \mathbf{v_i}) \mathbf{v_i}^T$$

$$A\mathbf{v_i} = \sigma_i \mathbf{u_i} \ \ and \ \ A^T \mathbf{u_i} = \sigma_i \mathbf{v_i}.$$

Truncation

- 性质2: 对于rank不超过k的任意矩阵B:

$$\|A - A_k\|_F \leq \|A - B\|_F \qquad A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

- Idea： 首先A-B的“成分”大小就是A-B每一行行向量“成分”大小之和。考虑每一对应的A B行向量α, β，固定B行向量β方向后，β模长只有取α在β方向上投影大小时，$|\alpha - \beta|$才最小。

Truncation

- 性质2: 对于rank不超过k的任意矩阵B:

$$\|A - A_k\|_F \leq \|A - B\|_F \qquad A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

- 这个性质带来了两个SVD的应用：图片压缩、数据(比如高维向量)降维(即PCA)

Image Compression

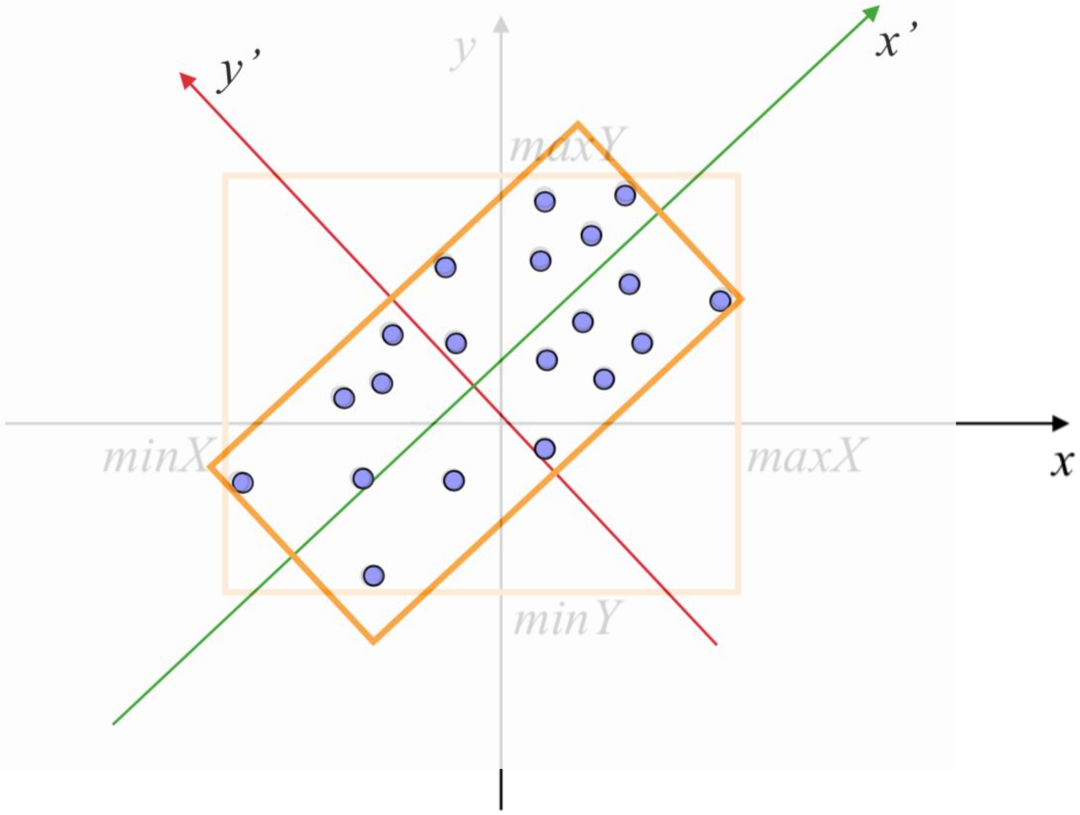
- 将上面的性质2，应用到图片压缩
- 如果存一张图 复杂度O(mn)
- 但如果能限制这张图到k-rank 而k << min{n,m}
- 存储的空间复杂度可以降到O(k * max(n,m))
- **(由性质2)而所有(不超过)k-rank矩阵中，SVD的k-部分和是最好的估计**

$$\|A - A_k\|_F \le \|A - B\|_F$$

$$A_k = \sum_{i=1}^{k} \sigma_i \mathbf{u_i} \mathbf{v_i}^T$$

PCA

给你一堆高维空间中的点，如何给它们找一组“好”的正交基？



We find better axes!

PCA

首先坐标系的中心应该是点云的重心
那我们先将所有的X_i中心化变成Y_i：

$$\mathbf{y}_i = \mathbf{x}_i - \mathbf{m}$$

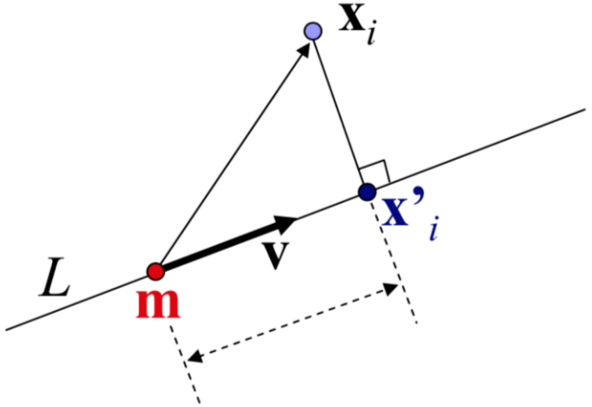
然后构造以下矩阵：

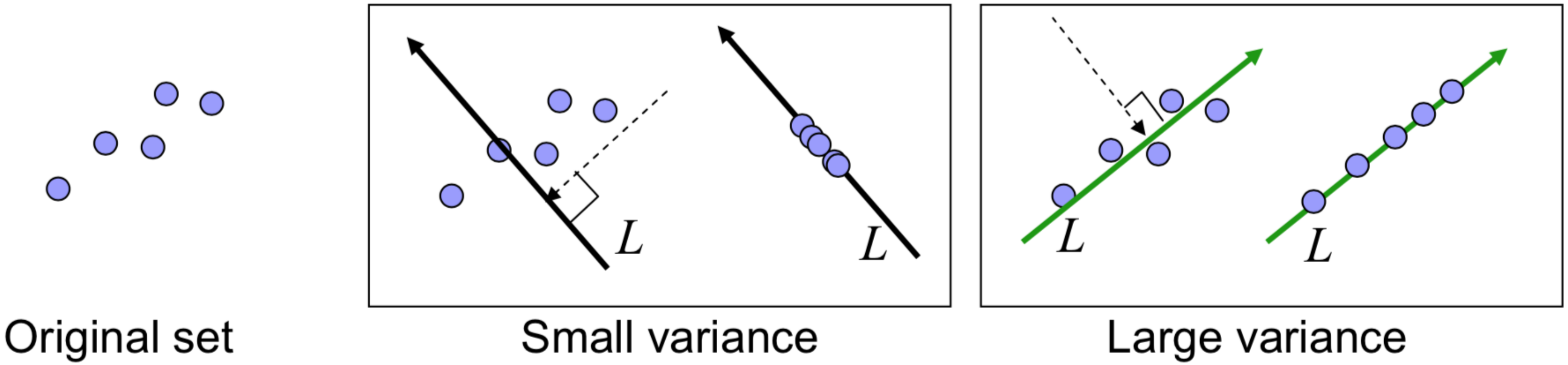
$$S = YY^\top \qquad Y = \begin{bmatrix} | & | & & | \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \\ | & | & & | \end{bmatrix}$$

PCA

考虑一个从原点出发的单位向量v，那么向量簇$\{y_1, y_2 \dots y_n\}$在v方向上的“成分”，刻画了向量簇v在方向的方差：

$$\mathrm{var}(L) = \frac{1}{n} \boxed{\sum_{i=1}^{n} \| \mathbf{x}'_i - \mathbf{m} \|^2}$$





PCA

把该式进一步展开

$$\mathrm{var}(L) = \frac{1}{n}\sum^{n} \| \mathbf{x}_i' - \mathbf{m} \|^2$$

$$\mathrm{var}(\ell) = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i' - \mathbf{m}\|^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}^\top\mathbf{y}_i)^2 = \frac{1}{n}\|\mathbf{v}^\top Y\|^2$$

$$= \boxed{\frac{1}{n}(\mathbf{v}^\top Y)(\mathbf{v}^\top Y)^\top = \frac{1}{n}\mathbf{v}^\top YY^\top\mathbf{v}} = \frac{1}{n}\langle S\mathbf{v}, \mathbf{v}\rangle\,.$$

矩阵YY^T究竟在刻画什么？

Evaluation Matrix

这是个度量阵，刻画的是对单位向量v, 向量簇$\{y_1, y_2 \dots y_n\}$在v上**成分的大小**！

$$\operatorname{var}(\ell) = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i' - \mathbf{m}\|^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}^\top \mathbf{y}_i)^2 = \frac{1}{n}\|\mathbf{v}^\top Y\|^2$$

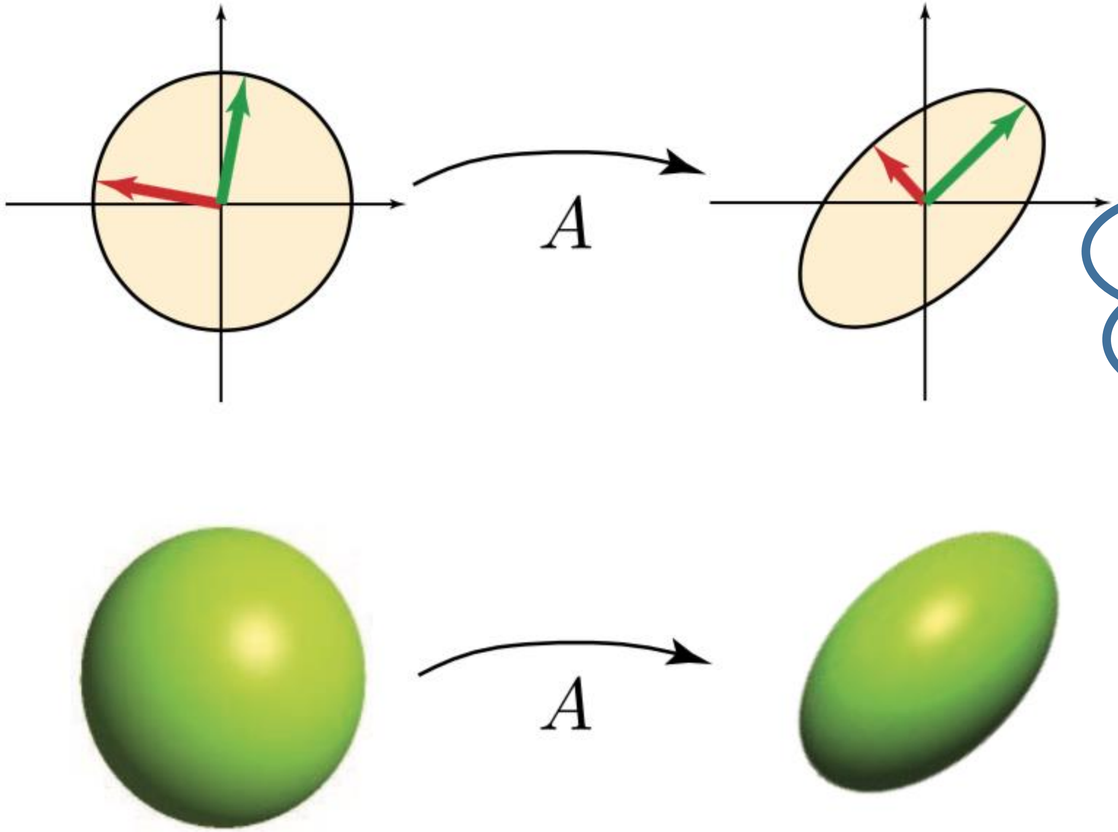
$$= \boxed{\frac{1}{n}(\mathbf{v}^\top Y)(\mathbf{v}^\top Y)^\top = \frac{1}{n}\mathbf{v}^\top Y Y^\top \mathbf{v}} = \frac{1}{n}\langle S\mathbf{v}, \mathbf{v}\rangle .$$

Two Perspective

- 我们从两种不同的视角分析SVD
- 一种是John的，将矩阵A视为向量簇
- 另一种是将矩阵A看成一个线性映射
- 咱继续从线性映射的角度观察

Linear Mapping

- 所有的线性映射 基本都长这样：



先“转”下坐标轴，
再改变模长？

Linear Mapping

- 如果映射**A**是可对角化的，那就存在一组基使得A在这组基下的矩阵是对角
- 但对general case，如果A是不可对角化的方阵？甚至是m * n?
- (加条件)一组基不够，再来一组：

$$\mathbf{A} \;:\; \mathbf{U} \rightarrow \mathbf{V}$$

$$\mathrm{A} \qquad (\alpha_1, \dots, \alpha_n) \qquad (\beta_1, \dots, \beta_m)$$

$$\mathbf{A}(\alpha_1, \dots, \alpha_n) = (\beta_1, \dots, \beta_m)\,\mathrm{A}$$

Linear Mapping

- 当基U V变化时，映射A对应的矩阵也随之改变，什么时候最简单？能否化成对角型？

设 $(\alpha_1', \dots, \alpha_n') = (\alpha_1, \dots, \alpha_n) Q$,

$(\beta_1', \dots, \beta_m') = (\beta_1, \dots, \beta_m) P$,

P, Q 是正交矩阵，则

$$\mathbf{A}(\alpha_1', \dots, \alpha_n') = [\mathbf{A}(\alpha_1, \dots, \alpha_n)] Q$$

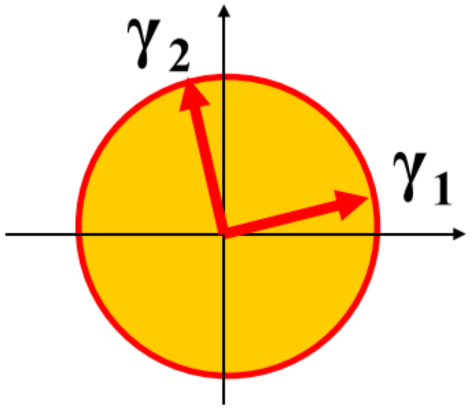
$$= (\beta_1, \dots, \beta_m)[A \ \ Q]$$

$$= (\beta_1', \dots, \beta_m')[P^T \ A \ \ Q]$$

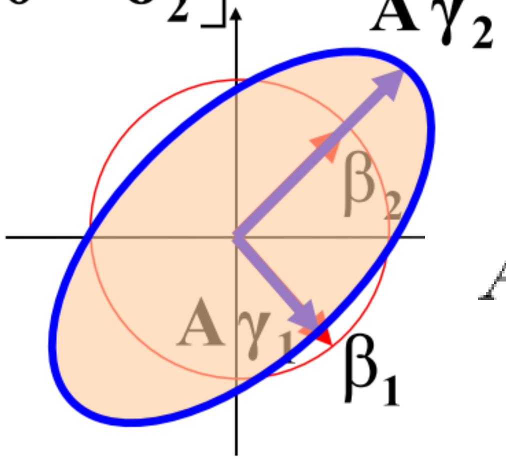
Linear Mapping

- 有了SVD 一目了然

$$A[\gamma_1\ \gamma_2] = [\beta_1\ \beta_2]\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$



A^TA特征向量

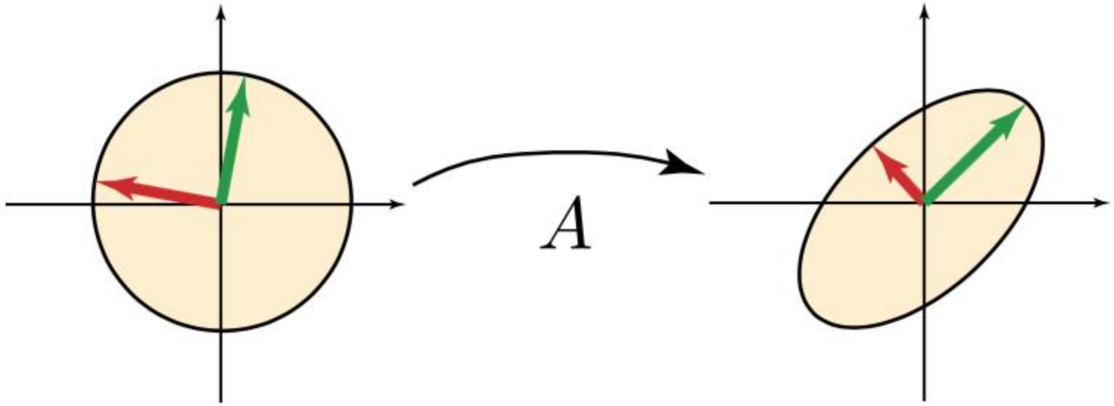


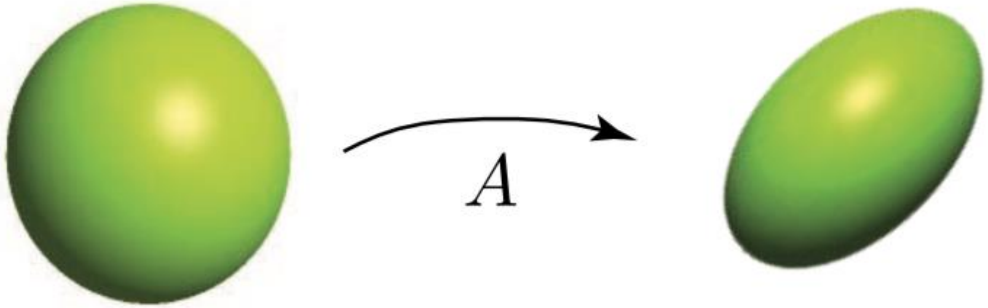
AA^T 特征向量

$$A\mathbf{v_i} = \sigma_i\mathbf{u_i}\ \ and\ \ A^T\mathbf{u_i} = \sigma_i\mathbf{v_i}.$$

Linear Mapping

- 回过头来再看最开始提到的linear mapping性质：





定理: $A = PSQ^T = (PQ^T)QSQ^T$

欧氏空间上的线性变换都可唯一地写成一个正交变换与一个对称变换的乘积.

Understanding Principal Component

- SVD也叫主成分分解，式子中的$\mathrm{u}_i v_i^T$正是不同的“主成分”
- 怎么(不严格地)理解主成分呢？
- 还是两种角度 把A看成行向量组合 、 或看成一个线性映射？

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_\mathbf{i} \mathbf{v}_\mathbf{i}^T.$$

- 仅Make sense一下，就不写slides上了

Other Matrix Decomposition

- SVD的分解是和矩阵自身相关的
- 能不能直接找一组基，independent from matrix 同时又有很强的解释性和应用呢？
- Spectral Decomposition

Other Matrix Decomposition

- Spectral Decomposition
- (就是离散傅里叶)

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \cos\frac{\pi}{16} & \cos\frac{2\pi}{16} & \cdots & \cos\frac{7\pi}{16} \\ \frac{1}{\sqrt{2}} & \cos\frac{3\pi}{16} & \cos\frac{6\pi}{16} & \cdots & \cos\frac{3\times 7\pi}{16} \\ & & & \cdots & \\ \frac{1}{\sqrt{2}} & \cos\frac{15\pi}{16} & \cos\frac{2\times 15\pi}{16} & \cdots & \cos\frac{7\times 15\pi}{16} \end{bmatrix}_{8\times 8}$$

列向量由低频到高频排列

$\{\alpha_i \alpha_j^T\}$构成线性空间$M_{8,8}(R)$的基底

X = 

线性表出的系数如何计算?

$\alpha_i \alpha_j^T\ (0 \le i, j \le 7)$的图像