先简单回顾下主成分析 `PCA(principle component analysis)` 与奇异值分解 `SVD(singular value decomposition)` 。

一、主成分析PCA

1、所解决问题

$y_{k\times1}=P_{k\times n}\cdot x_{n\times1}$

给定 m 个 n 维样本 $X=\{x_0,x_1,\ldots,x_m\}$ ，通过变换 $y=Px$ (其中 $P_{k\times n}$ 为变换矩阵)，将样本 $(x_i)_{i=0,\ldots,m}$ 从 n 维降到 k 维 $(y_i)_{i=0,\ldots,m}$ ，计 $Y=\{y_0,y_1,\ldots,y_m\}$ ，同时最大程度的减少降维带来的信息损失。

2、所依赖的原则

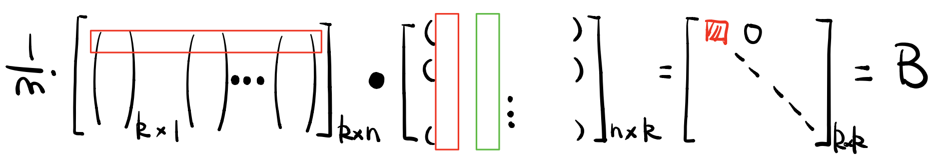
根据降维并减小信息损失的目标，可以得出以下两个原则

→基于标准正交基 $W^T\cdot W=I$ 的投影

1. 降维后的各个维度之间相互独立，即去除降维之前样本 x 中各个维度之间的相关性。
2. 最大程度保持降维后的每个维度数据的多样性，即最大化每个维度内的方差

记降维后样本集 Y 的协方差矩阵为

$$B_{k\times k}=\frac{1}{m}YY^T$$



上述第一个条件要求协方差矩阵B除了对角线上元素外，其他均为0，也即 B 为对角矩阵。

不同基的投影之间线性无关，正交.

将变换关系 $y=Px$ 代入Y的协方差矩阵B中，

$$B_{k\times k}=\frac{1}{m}YY^T=\frac{1}{m}PX(PX)^T=P\frac{1}{m}XX^TP^T=P_{k\times n}C_{n\times n}P^T_{n\times k}$$ --------(1)

其中， $C_{n\times n}=\frac{1}{m}XX^T$ 是变换前数据 X 的协方差矩阵。

$C_{n\times n}$ 的特征值分解形式如下：

正交

$$D_{n\times n}=Q_{n\times n}C_{n\times n}Q^T_{n\times n}$$ --------(2)

其中， $D_{n\times n}$ 为对角矩阵。

明显的式(1)和式(2)除了维度不同，其他均一样。

结合上述第二条原则，变换矩阵 $P_{k\times n}$ 即是矩阵C的前k大的特征向量按行组成的矩阵。

3、问题求解方法

式2就是协方差矩阵 C 的特征值分解，变换矩阵 $P_{k\times n}$ 即是矩阵C的前k大的特征向量按行组成的矩阵。所以，PCA的求解步骤为：

- 求 X 均值
- 将 X 减去均值
- 计算协方差矩阵 $C=\frac{1}{m}XX^T$
- 对协方差矩阵 C 特征值分解
- 从大到小排列 C 的特征值
- 取前 k 个特征值对应的特征向量按行组成矩阵即为变换矩阵 $P_{k\times n}$

这里的核心问题是协方差矩阵 $C=\frac{1}{m}XX^T$ 的特征值分解

如果理解了PCA笔记，这一节很好懂，就是对PCA的一种理解方式，其实只看最后一句即可。

二、奇异值分解SVD

1、所解决问题

$$A_{m\times n}=U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$ --------(3)

其中 $U_{m\times m}$ 和 $V_{n\times n}$ 均为正交矩阵， $\Sigma_{m\times n}$ 为对角矩阵

奇异值分解要解决的问题是将 $A_{m\times n}$ 矩阵分解为对角矩阵 $\Sigma_{m\times n}$ ， $\Sigma_{m\times n}$ 中对角元素 σ_i 称为矩阵 $A_{m\times n}$ 的奇异值

2、问题求解方法

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

所以 V 是 A^TA 特征值分解的特征向量按列组成的正交矩阵， Σ^2 是 A^TA 特征值组成的对角矩阵，也可以看出 $A_{m\times n}$ 的奇异值 σ_i 是 A^TA 特征值 λ_i 的平方根。

$$\sigma_i=\sqrt{\lambda_i}$$ --------(4)

假如 A^TA 的特征向量为 v_i ，则根据式3和式4， U 中对应的 u_i 则可以由下式求出：

$$u_i=\frac{Av_i}{\sigma_i}$$ --------(5)

也即奇异值分解的关键在于对 A^TA 进行特征值分解。

如果理解了线性代数笔记中SVD部分，这一节只看这一句就行

三、PCA与SVD的关系

由上述分析可知，

1. PCA求解关键在于求解协方差矩阵 $C=\frac{1}{m}XX^T$ 的特征值分解
2. SVD关键在于 A^TA 的特征值分解。

很明显二者所解决的问题非常相似，都是对一个实对称矩阵进行特征值分解，

如果取：

$$A=\frac{X^T}{\sqrt{m}}$$

$X \to \frac{1}{m}XX^T$ 特征值分解 → 取特征值最大的k个特征向量构成 PCA变换矩阵 $(\;)_{k\times n}$

$X \to \frac{X^T}{\sqrt{m}}$ 奇异值分解 → 取 Σ^2 中最大的k个特征值对应的右奇异矩阵V中的k个向量，构成 PCA变换矩阵 $(\;)_{k\times n}$

则有：

$$A^TA=\left(\frac{X^T}{\sqrt{m}}\right)^T\frac{X^T}{\sqrt{m}}=\frac{1}{m}XX^T$$

SVD与PCA等价，所以PCA问题可以转化为SVD问题求解，那转化为SVD问题有什么好处？

有三点：总的来说，就是降低计算复杂度.

1. 一般 X 的维度很高， A^TA 的计算量很大
2. 方阵的特征值分解计算效率不高
3. SVD除了特征值分解这种求解方式外，还有更高效且更准确的迭代求解法，避免了 A^TA 的计算

其实，PCA只与SVD的右奇异向量的压缩效果相同。如公式图所示，即 $V^T_{n\times n}=(\;)_{n\times n}$

1. 如果取 V 的前 k 行(列)作为变换矩阵 $P_{k\times n}$ ，则 $Y_{k\times m}=P_{k\times n}X_{n\times m}$ ，起到压缩行即降维的效果
2. 如果取 U 的前 d 行(列)作为变换矩阵 $P_{d\times m}$ ，则 $Y_{n\times d}=X_{n\times m}P_{m\times d}$ ，起到压缩列即去除冗余样本的效果。这不是PCA干的.