

An Introduction to PCA

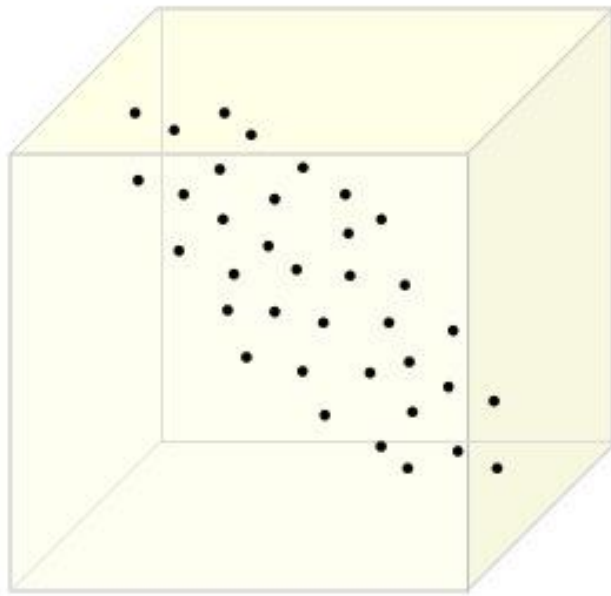
Praneeth Netrapalli

Microsoft Research New England

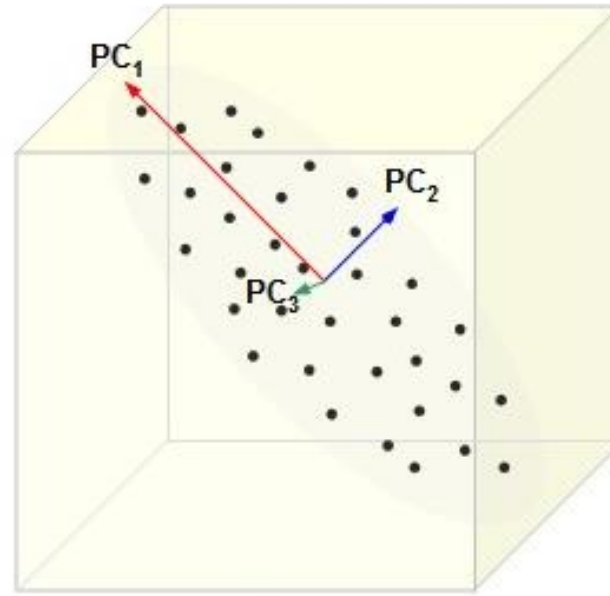
MSR Summer School on Machine Learning, IISc Bangalore

June 2015

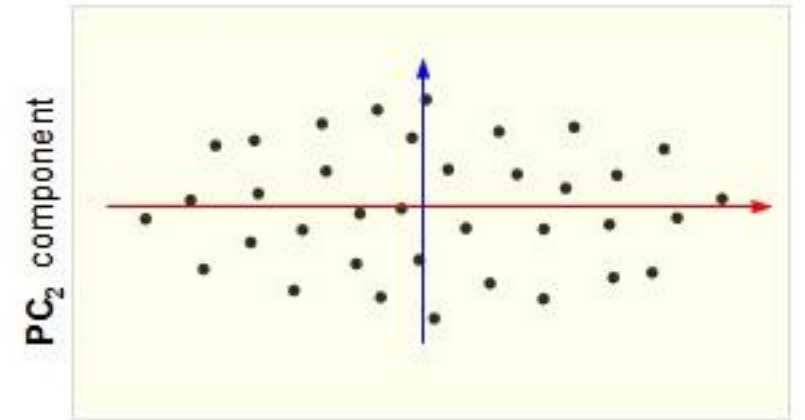
Dimensionality reduction



a



b

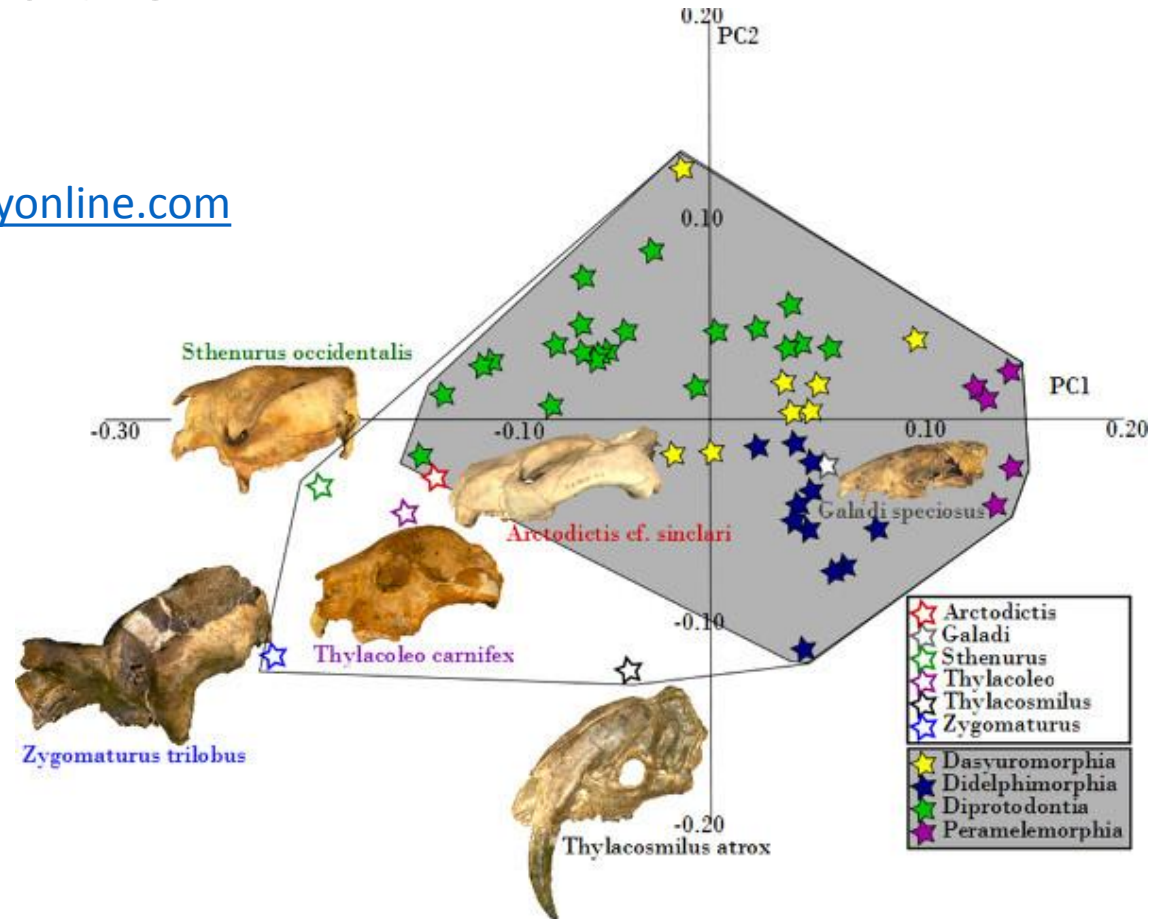


c

An application

Picture taken from

<http://www.palaeontologyonline.com>



Many more applications: solving system of linear equations, computing pseudoinverse, building block of many other algorithms...

Singular Value Decomposition (SVD)

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

- U, V – orthonormal column matrices
- Σ – positive diagonal matrix
- $\sigma_1 \geq \dots \geq \sigma_m$ - singular values
- Intuitive repsn. of transformation; unique
- For symmetric matrices, $U = \pm V$
- U eigenvectors of MM^T

Construction of SVD

$$\bullet v_1 = \operatorname{argmax}_{v: \|v\|_2=1} \|Mv\|_2$$

$$\bullet u_1 = \frac{Mv_1}{\|Mv_1\|_2}$$

$$\bullet \sigma_1 = \|Mv_1\|_2$$

$$v_{i+1} = \operatorname{argmax}_{\substack{v: \|v\|_2=1 \\ v \perp v_1, \dots, v_i}} \|Mv\|_2$$

$$u_{i+1} = \frac{Mv_{i+1}}{\|Mv_{i+1}\|_2}$$

$$\sigma_{i+1} = \|Mv_{i+1}\|_2$$

Relation to various norms

- Spectral norm/Operator norm : $\|M\|_2 := \max_{v:\|v\|_2=1} \|Mv\|_2 = \sigma_1$
- Frobenius norm : $\|M\|_F := \sqrt{(\sum_{i,j} M_{ij}^2)} = \sqrt{\sum_i \sigma_i^2}$
- Schatten norms : $\|M\|_{S_p} := (\sum_i \sigma_i^p)^{1/p}$

A simple application

Lemma: If M is square and has orthonormal columns then it also has orthonormal rows.

Proof:

$$\|Mv\|_2 = 1 \quad \forall v \in \mathbb{R}^n$$

$$\Rightarrow \sigma_i = 1 \quad \forall i$$

$$\Rightarrow \|u^T M\|_2 = 1 \quad \forall u \in \mathbb{R}^n$$

\Rightarrow Rows of M are orthonormal.

Eckart-Young-Mirsky Theorem

Let $M = U\Sigma V^T$ be the SVD of M . Then, for any rank- k matrix A , we have:

$$\|M - A\|_2 \geq \|M - U_k \Sigma_k V_k^T\|_2 = \sigma_{k+1}$$

$$\|M - A\|_F \geq \|M - U_k \Sigma_k V_k^T\|_F = \sqrt{\sum_{i=k+1}^m \sigma_i^2}$$

Proof of Eckart-Young-Mirsky Theorem

- Since $\text{rank}(A) \leq k$, $\dim(\text{null}(A)) \geq n - k$.
- Since $\dim(\text{null}(A)) + \dim(\text{range}(V_{k+1})) \geq n - k + k + 1 = n + 1$, there exists $x \in \text{null}(A) \cap \text{range}(V_{k+1})$.
- $\|(M - A)x\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 \langle x, v_i \rangle^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \langle x, v_i \rangle^2 = \sigma_{k+1}^2$

Proof of Eckart-Young-Mirsky Theorem

Frobenius norm for rank-1 case

$$\min_{u,v} \|M - uv^T\|_F^2$$

Stationary points:

$$\begin{aligned}(M - uv^T)v &= 0 \\ u^T(M - uv^T) &= 0\end{aligned}$$

$\Rightarrow u, v$ are singular vectors of M

Optimality $\Rightarrow u, v$ are top singular vectors!

Perturbation questions

It is often the case that data is corrupted, and we observe

$$M = L + N$$

where L is the desired matrix and N is noise.

Would like to answer questions such as

- How large is $(\sigma_k(M) - \sigma_k(L))$?
- How far are $u_k(M)$ and $u_k(L)$?

Perturbation inequalities

$$M = L + N$$

- Weyl's inequalities

$$|\sigma_k(M) - \sigma_k(L)| \leq \|N\|_2$$

- Davis Kahan theorem

$$\left\| (U_k(M))_{\perp}^T U_k(L) \right\|_2 \leq \frac{\|N\|_2}{|\sigma_k(L) - \sigma_{k+1}(L)|}$$

An exercise

- Suppose L is a rank- k matrix and $M = L + N$. Then,

$$\|L - P_k(M)\|_2 \leq 2\|N\|_2.$$

Proof: $\|L - P_k(M)\|_2 \leq \|M - L\|_2 + \|M - P_k(M)\|_2$

$$\leq 2\|M - L\|_2 \leq 2\|N\|_2.$$

A similar argument gives $\|L - P_k(M)\|_F \leq 2\|N\|_F$.

Another exercise

- Suppose L is a rank- k matrix and $M = L + N$. Then,

$$\|L - P_k(M)\|_F \leq 2\sqrt{2k} \|N\|_2.$$

Proof: Let U be the (at most $2k$) left singular vectors of $L - P_k(M)$.

$$\|L - P_k(M)\|_F^2 = \text{Tr} \left(U^T (L - P_k(M)) (L - P_k(M))^T U \right)$$

Write $L - P_k(M) = L - M + M - P_k(M)$, and use

$$\text{Tr}(U^T AB^T U) \leq 2k \|A\|_2 \|B\|_2$$

Summary

- SVD/PCA is a fundamental tool in many settings
- Perturbation analysis is crucial for these purposes
- Tight results known for perturbation in $\|\cdot\|_2$ and $\|\cdot\|_F$ norms
- Coming up later: perturbation bounds in $\|\cdot\|_\infty$ norm

References

- Linear Algebra Done Right, by Sheldon Axler
- Matrix Computations, by Gene Golub and Charles Van Loan
- <https://terrytao.wordpress.com/2010/01/12/254a-notes-3a-eigenvalues-and-sums-of-hermitian-matrices/>