SVD and Applications

(Singular Value Decomposition)

Girish Varma

Linear Algebra Recap

1. $\mathbb{R}^d$: all d-dimensional real vectors.
2. $B = \{b_1, b_2, \dots, b_d\}$ are a **basis** iff they are **linearly independent**.
    a. Example : [1,0], [0,1] for d = 2 called **standard basis**, denoted by e_1, e_2.
    b. Example : [1,1], [1,-1] for d = 2.
3. For any vector x in $\mathbb{R}^d$, and a basis $B = \{b_1, b_2, \dots, b_d\}$, there are **unique** $\alpha_1, .., \alpha_d \in \mathbb{R}$ such that $x = \alpha_1 b_1 + \dots + \alpha_d b_d$.
    a. They are called the **coordinates** of x with respect to the basis B
4. Let $\alpha = [\alpha_1, \dots, \alpha_d]$ and $\beta = [\beta_1, \dots, \beta_d]$ are coordinates with respect to basis B and C. Then there is a **change of basis matrix** M such that $\beta = M\alpha$.

Linear Algebra Recap 2

1. **Rank** of a matrix, **Norm** of a vector $|v|$
2. **Dot product** $x \cdot y = \sum x_i y_i$.
    a. If y is a unit vector, it is the **projection** of x along direction y.
3. $B = \{b_1, b_2, \dots , b_d\}$ are **orthogonal** if all pairwise dot products are 0. They are **orthonormal** if they are unit vectors and orthogonal.
4. For a matrix M, vector v and real λ, if $Mv = \lambda v$. Then λ is an **eigenvalue** and v an **eigenvector** of M.

Spectral Decomposition Recap

Let M be a **symmetric** matrix ($M^T = M$). Then there is an orthonormal **eigenbasis** $B = [b_1, b_2, \dots, b_d]$ and $\lambda_1, \dots, \lambda_d$ such that

$$M = B^T \operatorname{diag}(\lambda_1, \dots, \lambda_d)\, B.$$

Application: If a vector is expressed in the eigenbasis B of M, then computing Mx takes only d steps of computation (instead of d^2 steps of matrix multiplication).

SVD : Singular Value Decomposition

For any matrix M ϵ $\mathbb{R}^{n \times d}$ can be decomposed into M = U D V^T where columns of U ϵ $\mathbb{R}^{n \times n}$, V ϵ $\mathbb{R}^{d \times d}$ are **orthonormal** and D ϵ $\mathbb{R}^{n \times d}$ is a diagonal matrix with **positive** real entries.

Columns of V are called **left singular vectors** (and U **right singular vectors**) and diagonal entries of D denoted by $s_1, \ldots, s_r$ the **singular values**.

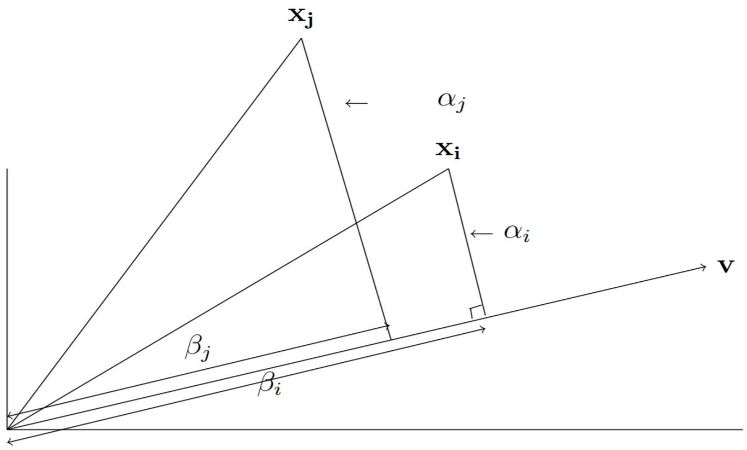
$$M(\alpha_1 v_1 + \ldots + \alpha_d v_d) = s_1 \alpha_1 u_1 + \ldots + s_d \alpha_d u_d$$

Pros : Defined for non square, non symmetric matrices also.

Best Least Squares Fit

Let $X^1, \ldots, X^n \in \mathbb{R}^d$ (n d-dim points).
Goal : Find a **1-dim subspace** (line passing through origin) which is the best fit.
best fit means one which minimizes sum of squares of perp. distances.

(dist of point to line)^2 = $|X^1|^2$ - (len of proj)^2



Min $\sum \alpha^2$

Equivalent to

Max $\sum \beta^2$

Singular Vectors

Consider the matrix $A \in \mathbb{R}^{n \times d}$ whose rows are $X^1, \ldots, X^n$. Let $v \in \mathbb{R}^d$ be the unit vector along the line. Then $X^1 \cdot v$ gives the projection on the line.

First singular vector v_1 of A is defined as and first singular value is $|Av_1|$

$$\mathbf{v_1} = \arg\max_{|\mathbf{v}|=1} |A\mathbf{v}|.$$

Best Fit 2-dim subspace:

The second singular vector v_2 of A is defined as

$$\mathbf{v_2} = \arg\max_{\mathbf{v}\perp\mathbf{v_1}, |\mathbf{v}|=1} |A\mathbf{v}|.$$

$$\mathbf{v_3} = \arg\max_{\mathbf{v}\perp\mathbf{v_1}, \mathbf{v_2}, |\mathbf{v}|=1} |A\mathbf{v}|$$

Example

Take many observations

SVD will give a single non zero singular value

Singular vector will be along the direction x, in the 6D space.

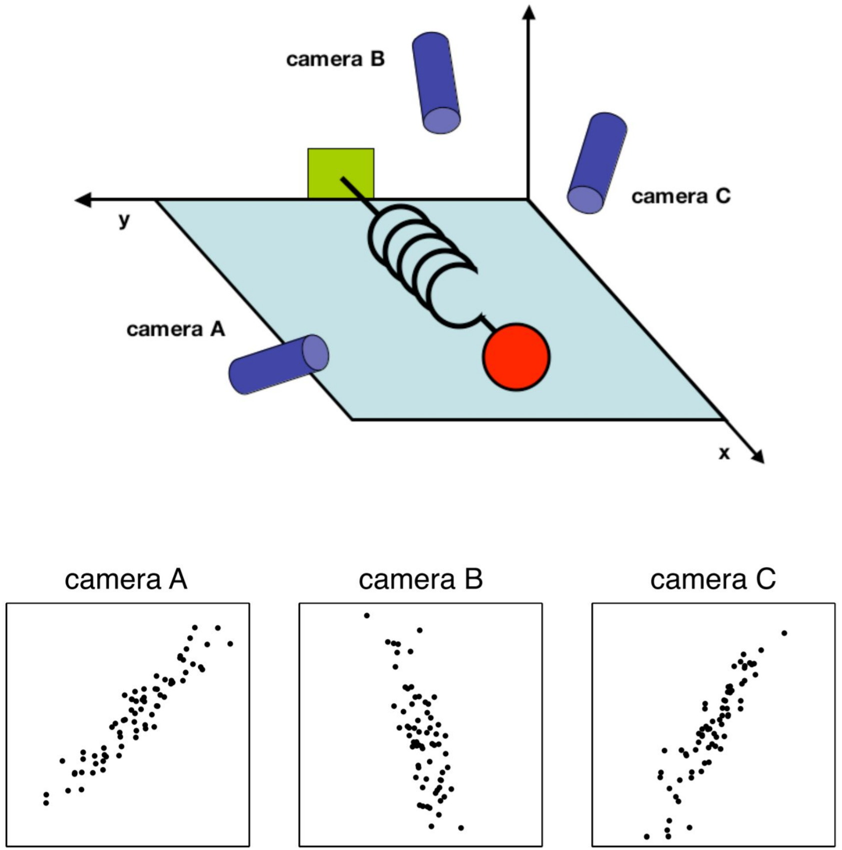
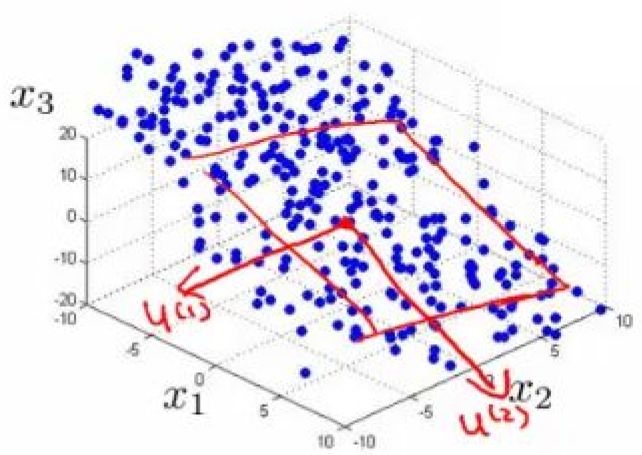


FIG. 1 A toy example. The position of a ball attached to an oscillating spring is recorded using three cameras A, B and C. The position of the ball tracked by each camera is depicted in each panel below.

$$\vec{X} = \begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \\ x_C \\ y_C \end{bmatrix}$$

Compression (Dimensionality Reduction)

Example : Customer - Product Data
n customers buying d products
Matrix A = (a_{ij}) a_{ij} is the prob. that i buys j



Hypothesis: customer purchase behaviour depends only on k underlying factors like age, income, family size etc. k << d

Then A = UV where U ϵ $\mathbb{R}^{n \times k}$ and V ϵ $\mathbb{R}^{k \times d}$.

nxk + kxd <<< nxd

Document Retrieval : Latent Semantic Analysis

Given a document q, get a **ranked** list of **similar** documents in your database.

Term-Document Matrix: n documents having d important terms. Represent each document as a d-dim vector which has the counts of the term in the document.

To get ranked list: find the term-document vector for query q and take dot products with vectors in database. Rank according to the dot products.

Problem: d might be too large. But most vectors are sparse

Solution: Take SVD of the Term-Document Matrix and ignore singular values which are too small.

PCA : Principal Component Analysis

If we have n points $(X^1, \dots, X^n)$ on d dimensional space, we did SVD on the nxd matrix. This can take a long time when n is large.

Mean Subtraction: Let $\mathbf{X} = (\sum X^i)/n$. Let $Y^i = X^i - \mathbf{X}$

Covariance Matrix : $C = (c_{ij})$ is a **dxd dim symmetric** matrix.
$c_{ij} = (\sum_k Y^k_i Y^k_j)/n$

Check: SVD of symmetric matrix same as Spectral Decomposition.

Ignore the singular vectors corresponding to small singular values. Represent $(X^1, \dots, X^n)$ as coordinates along these few vectors.

References

For theorems & proofs : Book by John Hopcroft and Ravi Kannan
https://www.cs.cmu.edu/~venkatg/teaching/CStheory-infoage/book-chapter-4.pdf

A Tutorial on PCA by Jonathon Shlens
https://arxiv.org/pdf/1404.1100.pdf

Extra Topics

1. Computing SVD and Spectral Decomposition
2. Prove Best Fit k-dim Subspace Theorem