Singular value decomposition.

Problem :-

- n columns of data, each column in $\mathbb{R}^d$.
- Side information :- The "true" dimension of the data points is $k \ll d$, but the data measurement is noisy.

Example :- Let $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ be n unknown unit vectors in $\mathbb{R}^k$ and let M be a fixed but unknown matrix of dimension $d \times k$. Then the data matrix

$$A = \underbrace{M}_{d \times k} \cdot \underbrace{\begin{pmatrix} \vdots & \vdots & & \vdots \\ X^{(1)} & X^{(2)} & \cdots & X^{(n)} \\ \vdots & \vdots & & \vdots \end{pmatrix}}_{k \times n} + \text{noise}$$

Question :- Estimate M from observations of A.

One possible solution concept:-

- Without the noise term, A would be a rank k matrix.

Thus

→ Find the rank k matrix that is "closest" to A.

Frobenius norm

$$\|A\|_F^2 = \sum_{i,j} |A_{ij}|^2 = \mathrm{Tr}(AA^*)$$

The case $k = 1$:-

We want to find a unit vector $v \in \mathbb{R}^n$ and constants $c_1, \dots, c_d$ s.t.

$$d(A, \mathrm{span}(v))^2 = \left\| A - \begin{pmatrix} c_1 v^T \\ c_2 v^T \\ \vdots \\ c^d v^T \end{pmatrix} \right\|_F^2 \quad \text{is minimized.}$$

For the latter, we must choose

$$C_i = (A_{(i)}^T v)$$

when v is already given.

[$A_{(i)}$ is the ith row of A, seen as a column vector]

[Why?]

Then, we have

$$d(A, \mathrm{span}(v))^2 = \sum_{i=1}^{d} \| A_{(i)} - (A_{(i)}^T v) v \|^2$$

$$= \|A\|_F^2 - \sum_{i=1}^{d} |A_{(i)}^T v|^2$$

$$= \|A\|_F^2 - \|Av\|^2$$

Thus, to minimize $d(A, \mathrm{span}(v))$, we must choose

$$v^{(1)} = \arg\max_{\|v\|=1} \|Av\| .$$

Consider now the case of general k. We now want to find a vector space V_k and an $d \times n$ matrix B s.t.

s.t.

$$d(A, V_k)^2 = \min_{\substack{B \\ B_{(i)} \in V_k \ 1 \le i \le d}} \|A - B\|_F^2$$

is minimized (over the choice of V_k)

Now note that

$$\|A - B\|_F = \sum_{i=1}^{d} \|A_{(i)} - B_{(i)}\|_2^2 .$$

Thus given V_k, the minimizing B is one for which $B_{(i)}$ is the projection of $A_{(i)}$ onto V_k. Let P be this projection operator.

$$d(A, V_k)^2 = \sum \|A_{(i)} - P A_{(i)}\|_2^2$$

$$= \|A\|_F^2 - \sum_i A_{(i)}^T P A_{(i)} \quad — \text{②}$$

(Using $P = P^T$ and $P^2 = I$)

We now prove the following by induction.

$$V_k = \underset{\substack{V:\, \dim V = k \\ v^{(1)} \ldots v^{(k)} \text{ orthonormal} \\ \text{basis for } V}}{\arg\max} \sum_{j=1}^{k} \|A v^{(j)}\|_2^2$$

The claim for V_1 is already proved above. The claim for general V_k follows from (2) [Why?]

We therefore get a set of "singular vectors"

$$v^{(i)} = \underset{\substack{v:\ \|v\|=1 \\ v \perp v^{(1)} \ldots v^{(i-1)}}}{\arg\max} \|Av\| \qquad 1 \leq i \leq d.$$

In particular, v is the ith e.v. of $A^T A$, which is a symmetric PSD matrix.

Some properties of the SVD [Exercise]

(1) If $v^{(i)}$ are the singular vectors of A with singular values σ_i then

$$A = \sum_i \sigma_i \, u^{(i)} v^{(i)T},$$

where $u^{(i)}$ are orthonormal.

Thus if A is a $d \times n$ matrix with

Then A can be written [assuming $d < n$]

$$A = UDV^T$$

where D is a $d \times d$ diagonal matrix with

$$D_{ii} = \sigma_i, \quad 1 \leq i \leq d$$

U is a $d \times d$ matrix with the ith column

$$U^{(i)} = u^{(i)}, \quad 1 \leq i \leq d \quad \text{and}$$

V is a $n \times d$ matrix with

$$V^{(i)} = v^{(i)} \quad 1 \leq i \leq d.$$

Further,

$$U^T U = I_{d \times d} = UU^T.$$

and $$V^T V = I_{d \times d}.$$

(2) Although we derived the SVD as a method to provide the best rank-k approximation to a given matrix in the Frobenius norm, SVD also provides the

best rank-k approximation according to the operator norm.

Pf: Let B be an arbitrary rank-k matrix. Then $B = \sum_{i=1}^{k} c_i x^{(i)} y^{(i)T}$ where $y^{(i)}$ are orthonormal. Note that V_{k+1} must contain a unit vector $\vec{z}$ that is not in $\text{span}(y^{(1)}, \dots, y^{(k)})$, since V_{k+1} is of dimension $k+1$. Thus we have

$$\|(A-B)z\| = \|Az\| \geq \sigma_{k+1} \quad \text{since } \|z\| = 1 \text{ and } z \in V_{k+1}.$$

On the other hand, if $B = \sum_{i=1}^{k} \sigma_i u^{(i)} v^{(i)T}$, then $\|A-B\| = \left\| \sum_{i \geq k+1} \sigma_i u^{(i)} v^{(i)T} \right\| = \sigma_{k+1}$. ∎