9 Applications of the SVD

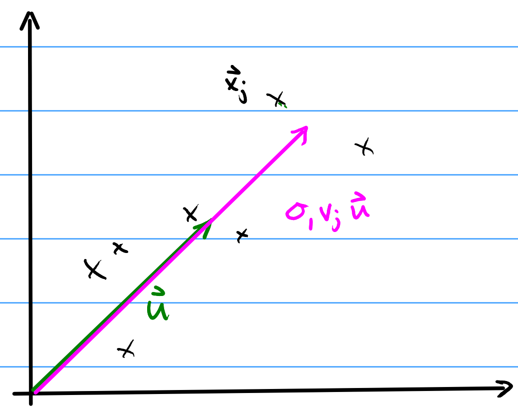
(1) Rank-k approximation

Let's start with the simplest case: rank-1 approximation, i.e. k = 1.

Suppose we've got a "matrix of points", i.e. a 2×n matrix X.

2 only for ease of illustration
works in higher dimensions, too

$$X = \begin{pmatrix} | & & | \\ \vec{x}_1 & \cdots & \vec{x}_n \\ | & & | \end{pmatrix} \qquad \vec{x}_i \in \mathbb{R}^2$$



Then find the reduced SVD:

$$\underbrace{U}_{2\times 2}\underbrace{\Sigma}_{2\times 2}\underbrace{V^T}_{2\times n} = X$$

Then find the rank-1 approximation: $X_1 = \vec{u}\,\sigma_1 \vec{v}^T$

(where $\vec{u}$ and $\vec{v}$ are the first columns of U, V)

And we know that $\|X - X_1\|_F^2$ is as small as it can be for any matrix of this form.

break down by columns

$$\|X - X_1\|_F^2 = \sum_j \|\vec{x}_j - \vec{u}\,\sigma_1 v_j\|_2^2$$

I.e. we have found a 2-vector $\vec{u}$ and n factors in a vector $\vec{v}$ so that the distances between the jth data point $\vec{x}_j$ and the multiples $\sigma_1 v_j \vec{u}$ of the vector $\vec{u}$ are minimal.

Demo: Rank-1 approximation

So now how about rank-k approximation?

Rank-2 approximation is analogous to the rank-1 case, except you find two vectors spanning a plane that has minimal distance from the data points.

Rank-k approximation is analogous to the rank-1 case, except you find k vectors spanning a plane that has minimal distance from the data points.

Give an example of where rank-k approximation does something useful.

Demo: Image Compression

(2) Computing the 2-norm

$$\|A\|_2 = \sigma_1$$

(3) Computing the 2-norm condition number

Assume A is invertible (and square). In this case:

$$\kappa(A) = \|A\| \|A^{-1}\| = \sigma_1 / \sigma_n$$

$$A = U \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} V^T$$

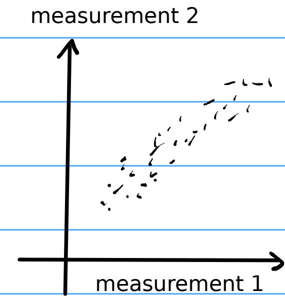
$$\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0$$

$$A^{-1} = V \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n} \end{pmatrix} U^T$$

If A is non-square, $\frac{\sigma_1}{\sigma_n}$ is still the 2-norm condition number of A.

In particular, if $\sigma_n = 0$, then the condition number is infinity.

(4) Principal Component Analysis ("PCA")



Have: a pile of "data"

More precisely: m 'measurements' from n 'trials'

each resulting in a real number

x_{ij} $\quad i = 1 \dots m \quad j = 1 \dots n$

Data matrix: $X = (x_{ij}) = \begin{pmatrix} | & | & | & | & | & | \end{pmatrix}$ ↓ i: measurements

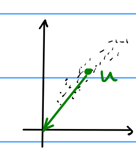
→ j: trials

Want: Underlying relationships

"If measurement i changes, the other measurements change along with it (in a computable manner) in most trials."

How do I compute a PCA?

(1) Compute an estimate of the means: $u_i = \frac{1}{n} \sum_j x_{ij}$



(2) Remove the means: $y_{ij} = x_{ij} - u_i$

↳ matrix Y

(3) Compute (an estimate of) the covariance matrix: $C_{i_1 i_2} = \frac{1}{n-1} \sum_j Y_{i_1 j} Y_{i_2 j}$

$\leadsto C = \frac{1}{n-1} Y Y^*$

n-1 to obtain unbiased estimate of covariance.

How do I compute a PCA? (cont'd)

Sums over trials, computes 'similarity' of pairs of measurements ('similarity' expressed as a dot product)

Observe: Large off-diagonal entries in C correspond to redundant measurements.

Idea: Find 'independent' measurements.

(4) Diagonalize the (s.p.d.) covariance matrix

$$\underbrace{\Sigma^2}_{\text{diag.}} = U C \underbrace{U^T}_{\text{orth.}}$$

(5) Transform Y to be 'independent'

Find V so that

$$\frac{1}{\sqrt{n-1}} Y = U \Sigma V^T$$

→ "Explained" measurements as linear combination of independent/"principal" components in the columns of U.

(6) Realize that this is the same calculation that led to the SVD.

→ All we need to do is compute an SVD of $\frac{1}{\sqrt{n-1}} Y$.

(5) Least squares for underdetermined and singular systems

Want to solve $Ax \cong b$ when A has a nullspace.

That is, there is a vector n≠0 so that An=0.

Suppose we have a solution x.

Then (x+αn) (for any scalar α) produces the same residual:

$$A(x+\alpha n) = Ax + \alpha An = Ax$$

$$\Rightarrow \|A(x+\alpha n) - b\|_2 = \|Ax-b\|_2$$

Demo: Solving least squares using the SVD (Part I)

The solution is not unique, which is a little sad.

We need more constraints than just minimizing the residual in order to get a unique solution.

Additional constraint: Minimize $\|Ax-b\|_2$ and $\|x\|_2$ simultaneously.

Can we use the SVD to solve the least-squares problem?

Now: Use the SVD $A = U\Sigma V^T$ to solve $Ax \cong b$.

$$\|Ax - b\|_2^2 = \|U\Sigma V^T x - b\|_2^2$$

$$= \|U^T(U\Sigma V^T x - b)\|_2^2$$

$$= \|\Sigma V^T x - U^T b\|_2^2 \qquad \leftarrow y = V^T x$$

$$= \|\Sigma y - V^T x\|_2^2$$

$$= \left\| \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_k & 0 \\ & 0 & & 0 \end{pmatrix} y - U^T b \right\|_2^2 \qquad \sigma_1, \dots \sigma_k \neq 0$$

minimized when $\sigma_1 y_1 = (U^T b)_1$

$\vdots$

$\sigma_k y_k = (U^T b)_k$

No conditions on $y_{k+1} \dots y_n$. $\leftarrow$ Set to zero.

Define $\Sigma^+ = \begin{pmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_k^{-1} & 0 \\ & 0 & & 0 \end{pmatrix}$. So $y = \Sigma^+ U^T b$.

Lastly, $V^T x = y = \Sigma^+ U^T b$. Solve: $x = V\Sigma^+ U^T b$.

This x minimizes the residual norm. It also minimizes $\|x\|_2$. $\rightarrow$ HW

$A^+ = V\Sigma^+ U^T$ is called the pseudo-inverse of A.

$x = A^+ b$ solves the least-squares problem uniquely for any A.

Demo: Solving least squares using the SVD (Part II)

(6) "Total" least squares

For a given matrix A, find the vector x so that

- $\|Ax\|_2$ is minimal
- $\|x\|_2 = 1$

Easy to find:

(1) Compute SVD of A : $A = U \Sigma V^T$

(2) Last column of V contains an (not necessarily the) answer