5374 Fall '22

Numerical Linear Algebra Lecture 23

More on the Singular Value Decomposition

Theorem (The SVD): Let A be a $m \times n$ matrix, then exist three matrices U, Σ, V such that:

U is $m \times m$, orthogonal

V is $n \times n$, orthogonal

Σ is $m \times n$, diagonal

and such that

$$A = U \Sigma V^t$$

Moreover, the numbers along the diagonal of Σ are all nonnegative and written as:

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \end{pmatrix} \quad (m \times n)$$

with $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$.

Remark: The numbers $\sigma_1, \sigma_2, \dots$ are uniquely determined, i.e. if $\exists$ a second triplet U', Σ', V' such that $A = U'\Sigma'(V')^*$ and U', V' are orthogonal, Σ' is diagonal with non-negative decreasing entries then $\Sigma = \Sigma'$. Thus the sequence of numbers $\sigma_1, \sigma_2, \ldots$ are a well defined property of A, and they are called the singular values of A. If A has a SVD decomposition with $A = U\Sigma V^*$, then the columns of U and V are called left and right eigenvectors (or singular vectors).

Proof/ Construction of the SVD

Consider A which is $m \times n$, and for now assume $m \geq n$ and $\operatorname{rank}(A) = n$, so in particular A is injective. Consider the function

$$f(x) = \|Ax\|_2 \qquad x \in \mathbb{R}^n$$

What is the gradient of $f(x)$?

Exercise: $$\nabla f(x) = \frac{A^* A x}{\|Ax\|_2} \qquad (x \neq 0)$$

(Hint: $f(x) = \sqrt{(Ax, Ax)} = \sqrt{(A^*Ax, x)}$)

Now suppose you want to find the largest and smallest values of $f(x)$ for x ranging over the unit sphere in $\mathbb{R}^n$. This is a constrained optimization problem that looks like this

$$\max_{x \in \mathbb{R}^n} f(x)$$
$$\text{subject to} \quad g(x) = 0$$

when $f(x) = \|Ax\|_2$ and $g(x) = \|x\|_2 - 1$

This problem can be solved via Lagrange multipliers, and what one finds is that at a maximum point x_0 of $f(x)$ in the set $\{g=0\}$, we must have

$$\nabla f(x_0) = \lambda \nabla g(x_0)$$

for some number λ. (This is called the Lagrange multiplier).

In this particular case, we would get

$$\frac{A^t A x_0}{\|Ax_0\|_2} = \lambda \frac{x_0}{\|x_0\|_2} = \lambda x_0 \quad \left(\begin{matrix}\text{since} \\ \|x_0\|_2 = 1\end{matrix}\right)$$

$$\Longleftrightarrow \qquad (A^t A) x_0 = \lambda \|Ax_0\|_2 \, x_0$$

This suggests that understanding how $f(x)$ varies goes through understanding the eigenvectors of A^tA.

Recall: A^tA is a positive-semi definite matrix (in fact, positive-definite since $\operatorname{rank}(A)=n$)

$$\left((A^tAx, x) = (Ax, Ax) = \|Ax\|_2^2 = f(x)^2 \right)$$

Then, we know from the theorem on diagonalization of symmetric matrices that there is a orthonormal basis $v_1, \dots, v_n$ made out of eigenvectors of A^tA; since A^tA is positive-definite, its eigenvalues are all positive, so they can be written as squares of positive numbers.

Shuffling the order of the v_k if necessary, we have

$$A^tA\, v_k = \sigma_k^2\, v_k \qquad \text{for } k = 1, \dots, n$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$.

Observe that

$$\begin{aligned} f(v_k) &= \sqrt{(Av_k, Av_k)} \\ &= \sqrt{(A^tAv_k, v_k)} \end{aligned}$$

$$= \sqrt{(\sigma_k^2 v_k, v_k)} = \sigma_k > 0$$

Then, let us define n vectors in $\mathbb{R}^m$ by

$$u_k := \frac{A v_k}{\sigma_k}, \quad \text{for } k = 1, \dots, n.$$

The u_k are unit vectors, plus

$$\begin{aligned}
(u_k, u_j) &= \left(\frac{A v_k}{\sigma_k}, \frac{A v_j}{\sigma_j} \right) \\
&= \frac{1}{\sigma_k \sigma_j} (A v_k, A v_j) \\
&= \frac{1}{\sigma_k \sigma_j} (A^t A v_k, v_j) \\
&= \frac{\sigma_k}{\sigma_j} (v_k, v_j) = \delta_{ij}
\end{aligned}$$

What do we have? We have:

- An orthonormal basis $v_1, \dots, v_n$ of $\mathbb{R}^n$
- n positive numbers $\sigma_1, \dots, \sigma_n$
- An orthonormal family $u_1, \dots, u_n$ of $\mathbb{R}^m$

where, for each $k = 1, \dots, n$

$$A v_k = \sigma_k u_k$$

In any way I like, let me pick an additional

$m-n$ vectors $u_{n+1}, u_{n+2}, \dots, u_m$ in $\mathbb{R}^m$ chosen such that $u_1, \dots, u_m$ is an orthonormal basis of $\mathbb{R}^m$. Then, write

$$V = \left(v_1 \mid v_2 \; \cdots \; \mid v_n \right) \qquad (n \times n)$$

$$\Sigma = m\left\{ \overbrace{\begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}^{n} \right. \qquad (m \times n)$$

$$U = \left(u_1 \mid u_2 \mid \cdots \mid u_m \right) \qquad (m \times m)$$

Clearly, given $x \in \mathbb{R}^n$

$$x = V V^t x$$

$$x = (x, v_1) v_1 + \cdots + (x, v_n) v_n$$

$$\Longrightarrow \quad Ax = (x, v_1) A v_1 + \cdots + (x, v_n) A v_n$$

$$= \sigma_1 (x, v_1) u_1 + \cdots + \sigma_n (x, v_n) u_n$$

$$= U \left(\Sigma V^t x \right)$$

$$= \left(U \Sigma V^t \right) x \qquad \forall \, x$$

So $$A = U \Sigma V^t$$

Remark : If $A = U \Sigma V^t$, then
$$A^t = V \Sigma^t U^t$$

For the case where A^tA is not invertible, we stop by defining $u_1, \dots, u_\ell$ when $\sigma_\ell \neq 0$ and $\sigma_{\ell+1} = 0$, and complete this orthonormal family to a full basis.

Remark: There is also the "reduced SVD" which takes the form:

$$A = \tilde{U} \tilde{\Sigma} \tilde{V}^t$$

where $\tilde{U}$ is $m \times n$, with orthonormal columns
$\tilde{\Sigma}$ is $n \times n$, diagonal, non-increasing and nonnegative diagonal entries
$\tilde{V}$ is $n \times n$, orthogonal.

Interesting facts about the SVD

1. If $A = U \Sigma V^t$, then

$$A^tA = (V \Sigma^t \cancel{U^t})(\cancel{U} \Sigma V^t)$$
$$= V (\Sigma^t \Sigma) V^t$$

Likewise,

$$AA^t = U \Sigma \Sigma^t U^t$$

2. Recall the Frobenius norm of a matrix

$$\|A\|_{Fr} = \sqrt{tr(A^t A)}$$

In light of the SVD,

$$\|A\|_{Fr} = \sqrt{tr(V \Sigma^t \Sigma V^t)}$$

V is orthogonal so $V(\Sigma^t \Sigma)V^t = V(\Sigma^t \Sigma)V^{-1}$

$$\Rightarrow \quad tr(V \Sigma^t \Sigma V^t) = tr(\Sigma^t \Sigma)$$

$$= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$$

In conclusion, $\|A\|_{Fr} = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$

3. The following characterization of the singular values and the singular vectors of A is crucial in numerical solvers for the SVD:

Let A be a $m \times n$ matrix. Consider

The following $(m+n)\times(m+n)$ matrix D

$$D = \begin{array}{cc} \begin{array}{cc} m & n \end{array} & \\ \left(\begin{array}{c|c} O & A \\ \hline A^t & O \end{array}\right) & \begin{array}{l} | \, m \\ | \, n \end{array} \end{array}$$

This is a symmetric $(m+n)\times(m+n)$ matrix.
Suppose $w \in \mathbb{R}^{m+n}$ is an eigenvector of D,

$$D w = \lambda w$$

Since $w \in \mathbb{R}^{m+n}$, we can write it as

$$w = \begin{pmatrix} u \\ v \end{pmatrix}, \text{ where } \begin{array}{l} u \in \mathbb{R}^m \\ v \in \mathbb{R}^n \end{array}$$

Then (check that the dimensions match)

$$D w = \begin{pmatrix} O & A \\ A^t & O \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}$$

$$= \begin{pmatrix} A v \\ A^t u \end{pmatrix} = \lambda w = \begin{pmatrix} \lambda u \\ \lambda v \end{pmatrix}$$

So w is an eigenvector $\Longleftrightarrow$

$$A v = \lambda u, \quad A^t u = \lambda v$$