INTRODUCTION: Some Basic Concepts from Linear Algebra

Throughout the class, all vector $u \in \mathbb{R}^m$ are column vectors. If A is a matrix, A_j, or a_j, or $A_{\star,j}$, or $A_{\cdot,j}$ denotes the j-th column of A. Similarly $A_{j,\star}$, or $A_{j,\cdot}$ will denote the j-th row vectors.

1. Matrix-vector product

Let $X \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^n$. $Xb \in \mathbb{R}^m$ is the usual product of X by b. It is often very useful to look at Xb as

$$Xb = \sum_{j=1}^{n} b_j X_j,$$

where X_j is the j-th column of X. Similarly if $Y = [y_1, \ldots, y_p] \in \mathbb{R}^{n\times p}$, we will often view the product XY as $[Xy_1, \ldots, Xy_p]$. In complete analogy, for $v \in \mathbb{R}^m$, we look at $v'X$ as

$$v'X = \sum_{k=1}^{m} v_k X_{k,\star},$$

as linear combination of rows of X.

Remark 1. It is easy to see that the cost (the number of basic arithmetic operations performed) in computing Xb is $2mn$. If $B \in \mathbb{R}^{n\times p}$, the cost of computing XB is $2mnp$. In particular if $X \in \mathbb{R}^{m\times m}$, the cost of computing X^2 is $2m^3$.

Here are some important but immediate consequence of this interpretation. The range of X is defined $\mathsf{Range}(X) \stackrel{\text{def}}{=} \{Xb,\ b \in \mathbb{R}^n\}$, also called the column space of X. From the above interpretation of Xb, we see that $\mathsf{Range}(X)$ is the space spanned by the columns of X, and we will sometimes use the notation $\mathsf{span}(X_{\cdot 1}, \ldots, X_{\cdot n})$. The null space of X is $\mathsf{Ker}(X) \stackrel{\text{def}}{=} \{b \in \mathbb{R}^n :\ Xb = 0\}$. Hence $\mathsf{Ker}(X) \neq \{0\}$ iif a non-trivial linear combination of the columns of X is zero, that is the columns of X are non linearly independent. The dimension of $\mathsf{Range}(X)$ is called the column-rank of X. And the row-rank of X is the dimension of the linear space spanned by the rows of X, that is $\mathsf{dim}\ \{v'X,\ v \in \mathbb{R}^m\}$. The column rank is always equal to the row-rank, and we simply speak of the rank of X and write $\mathsf{Rank}(X)$. Of course $\mathsf{Rank}(X) \leq m \wedge n$. An important theorem of linear algebra says that

$$\mathsf{Rank}(X) + \mathsf{dim}\ \mathsf{Ker}(X) = n.$$

2. Matrix norms

We introduce here the Frobenius norm and the spectral norm, two of the most commonly used matrix norm. The ℓ^p norm of a vector is $\|x\|_p = \{\sum_{i=1}^p |x_i|^p\}^{1/p}$, $p \geq 1$, with $\|x\|_\infty = \max_{1\leq i\leq p} |x_i|$. For a matrix $X \in \mathbb{R}^{m\times n}$ there are many norms. Although are all these norm are topologically equivalent, they have other different properties. One the most common matrix norm is the Frobenius norm. This is the norm we get by looking at X as a vector in $\mathbb{R}^{mn}$:

$$\|X\|_{\mathsf{F}} \stackrel{\text{def}}{=} \sqrt{\sum_{i,j} X_{ij}^2}.$$

The very nice property of the Frobenius norm is that the space $(\mathbb{R}^{m\times n}, \|\cdot\|_{\mathsf{F}})$ is a Euclidean space with inner product

$$\langle A, B\rangle = \mathsf{Tr}(AB') = \sum_{i,j} A_{ij}B_{ij}.$$

Another commonly used family of norms is the operator norms

$$\|X\|_{p,q} = \sup_{\|b\|_p\neq 0} \frac{\|Xb\|_q}{\|b\|_p}.$$

For $p = q$, we write $\|X\|_p$. Important special cases:

$$\|X\|_1 = \max_{1\leq j\leq n} \|X_{\star,j}\|_1,$$

$$\|X\|_\infty = \max_{1\leq i\leq m} \|X_{i,\star}\|_1.$$

For $p = q = 2$, $\|X\|_2$ is called the spectral norm of X. But unlike the case $p = 1, \infty$ presented above there is no easy formula for computing $\|X\|_2$. We will see below that $\|X\|_2$ is the largest singular value of X, and $\|X\|_2^2$ is the largest eigenvalue of $X'X$.

The Frobenius norm and the spectral norm are invariant under multiplication by a an orthogonal matrix: if $U \in \mathbb{R}^{n,n}$ and $V \in \mathbb{R}^{m\times m}$, are such that $U'U = I_n$, and $V'V = I_m$, then $\|VX\|_\star = \|XU\|_\star = \|X\|_\star$, for $\star \in \{2, \mathsf{F}\}$. It is also important to know that these two norm are sub-multiplicative. That is for $\star \in \{2, \mathsf{F}\}$, $Y \in \mathbb{R}^{n\times p}$,

$$\|XY\|_\star \leq \|X\|_\star \|Y\|_\star.$$

3. The Singular value decomposition and some applications

A very important generalization of the eigen-decomposition is the singular value decomposition (SVD). It is arguably the most fundamental matrix factorization. An important feature of the SVD it that it always exists.

A matrix $X \in \mathbb{R}^{m\times n}$ admits a SVD if there exist orthogonal matrix $U \in \mathbb{R}^{m\times m}$, orthogonal matrix $V \in \mathbb{R}^{n\times n}$ and diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ with non-negative diagonal elements in nonincreasing order such that $X = U\Sigma V'$. The diagonal elements of Σ are called singular values of X, the columns of U are called left-singular vectors and the columns of V are called right singular vectors. It is easy to check the the decomposition $X = U\Sigma V'$ can also be written as

$$X = \sum_{j=1}^{m\wedge n} \sigma_j u_j v_j',$$

where u_j (resp. v_j) is the j-th column of U (resp. V) and σ_j is the j-th singular value of X. Let

$$r \stackrel{\text{def}}{=} \max\{k \geq 1 : \ \sigma_k > 0\},$$

and set $r = 0$ if all $\sigma_j = 0$. Then the SVD of X can be written as

$$X = \sum_{j=1}^{r} \sigma_j u_j v_j'.$$

Using the svd of X we introduce the following important definition.

Definition 3.1. *The condition number of the matrix $X \in \mathbb{R}^{m\times n}$, is the quantity*

$$\kappa_2(X) \overset{\text{def}}{=} \frac{\sigma_1}{\sigma_{m\wedge n}} \in [1, +\infty].$$

Theorem 3.2. *Any matrix $X \in \mathbb{R}^{m\times n}$ admits a SVD. Furthermore, the singular values are unique.*

Proof. Since $X'X$ is symmetric semi-positive definite, all its eigenvalues are nonnegative real numbers. Let $\sigma_1^2 > \cdots > \sigma_r^2 > 0$ be the positive eigenvalues of $X'X$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$. Let $V = [v_1, \ldots, v_n]$ an orthogonal matrix of $\mathbb{R}^n$, made of eigenvectors of $X'X$. Set $V_1 = [v_1, \ldots, v_r]$, $V_2 = [v_{r+1}, \ldots, v_n]$, $S = \mathsf{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$, and $U_1 = XV_1S^{-1/2}$.
<u>Claim 1</u>: $U_1'U_1 = I_r$.

In view of Claim 1, we complete U_1 to form an orthogonal matrix of $\mathbb{R}^m$ $U = [U_1, U_2]$.

<u>Claim 2</u>: $U'XV = \begin{pmatrix} S^{1/2} & 0 \\ 0 & 0 \end{pmatrix}$.

Claim 1 holds because $X'XV_1 = V_1S$. FOr Claim 2, calculate that

$$U'XV = \begin{pmatrix} U_1' \\ U_2' \end{pmatrix} (XV_1 \;\; XV_2) = \begin{pmatrix} U_1'XV_1 & U_1'XV_2 \\ U_2'XV_1 & U_2'XV_2 \end{pmatrix}.$$

We have $(X'X)V_2 = 0$, so that $V_2'(X'X)V_2 = 0$. Hence $XV_2 = 0$. Also, $U_2'XV_1 = 0$ because $XV_1 = U_1S^{1/2}$. □

A SVD has many useful consequences.

Proposition 3.3. *Let $X \in \mathbb{R}^{m\times n}$ with SVD $X = \sum_{j=1}^r \sigma_j u_j v_j'$, with r as defined above.*

(1) $\mathsf{Range}(X) = \{Xb,\ b \in \mathbb{R}^n\} = \mathsf{span}(u_1, \ldots, u_r)$, $\{v'X,\ v \in \mathbb{R}^m\} = \mathsf{span}$(first r rows of V'), and $\mathsf{Ker}(X) = \mathsf{span}(v_{r+1}, \ldots, v_n)$.

(2) $\mathsf{Rank}(X) = r$, $\|X\|_2 = \sigma_1$ and $\|X\|_F = \sqrt{\sum_{k=1}^r \sigma_k^2}$.

(3) If $m = n$, $\det(X) = \pm\left(\prod_{i=1}^n \sigma_i\right)$.

Proof. (1) Follows readily from $X = \sum_{j=1}^r \sigma_j u_j v_j'$.

(2) Given the characterization of $\mathsf{Range}(X)$ given in (1), it follows automatically that $\mathsf{Rank}(X) = r$. For $\star \in \{2, \mathsf{F}\}$, $\|X\|_\star = \|U\Sigma V'\|_\star = \|\Sigma\|_\star$. Then calculate that $\|\Sigma\|_2 = \sigma_1$, and $\|\Sigma\|_{\mathsf{F}} = \sqrt{\sum_{k=1}^r \sigma_k^2}$.

(3) This is because $\det(U) = \pm 1$, and $\det(V) = \pm 1$.

□

Proposition 3.4. *The singular values of X are the square root of the eigenvalues of $X'X$ or XX'. Furthermore, if X is symmetric, the singular values of X are the absolute values of the eigenvalues of X.*

Proof. We easily check that $X'X = V\Sigma'\Sigma V'$, which is the eigen-decomposition of $X'X$. Similarly, $XX' = U\Sigma\Sigma' U'$. If X is symmetric, its eigen-decomposition is $X = W\Lambda W'$, with the diagonal matrix Λ in nonincreasing absolute value order. The SVD of X follows as $X = W\,[\Lambda \mathsf{sign}(\Lambda)]\,[\mathsf{sign}(\Lambda) W']$, where $\mathsf{sign}(\Lambda)$ is the diagonal matrix of signs of Λ. ☐

Theorem 3.5 (Eckart-Young theorem)**.** *For $q < r$, set $B_q = \sum_{j=1}^{q} \sigma_j u_j v_j'$. Then*

$$B_q \in \mathit{Argmin}_{\{B \in \mathbb{R}^{m\times n},\ \mathit{Rank}(B)=q\}} \|X - B\|_2.$$

The result continues to hold with the Frobenius norm.

Proof. Clearly B_q is of rank q, and $X - B_q = UDV'$ where $D = \mathsf{diag}(0,\ldots,0,\sigma_{q+1},\ldots,\sigma_r,0,\ldots,0)$. Hence $\|X - B_q\|_2 = \sigma_{q+1}$. Now take $B \in \mathbb{R}^{m\times n}$ such that $\mathsf{Rank}(B) = q$. Then $\mathsf{dimKer}(B) = n - q$. Since $n - q + q + 1 > n$, we can find $v_0 \in \mathsf{Ker}(B) \cap \mathsf{span}(v_1,\ldots,v_{q+1})$ with $\|v_0\| = 1$. Therefore

$$\|(X-B)v_0\|_2^2 = \|Xv_0\|_2^2 = \|U\Sigma V' v_0\|_2^2 = \|\Sigma V' v_0\|_2^2 = \sum_{j=1}^{q+1} \sigma_j^2 \left\langle v_j, v_0\right\rangle^2 \geq \sigma_{q+1}^2 \sum_{j=q+1}^{2} \left\langle v_j, v_0\right\rangle^2 = \sigma_{q+1}^2.$$

Hence for any matrix $B \in \mathbb{R}^{m\times n}$ such that $\mathsf{Rank}(B) = q$, $\|X - B\|_2 \geq \|X - B_q\|_2$. ☐