Mathematics of Machine Learning
Homework 2 - Solutions

Instructors: Afonso S. Bandeira & Nikita Zhivotovskiy
Course Coordinator: Pedro Abdalla Teixeira

March 12, 2021

Try to solve the questions before looking to the answers. Every item must be proved rigorously.

Problem 1

Prove Proposition 3.1 in the lecture notes.

Solution 1

- Let $X \in \mathbb{R}^{n\times m}$ be a matrix with r non-zero singular values. This means that X can be written as $X = U\Sigma V^T$ where $U \in \mathcal{O}(n)$, $V \in \mathcal{O}(m)$ and $\Sigma \in \mathbb{R}^{n\times m}$ is diagonal with only r non-zero values on the diagonal. It follows that the rank of Σ is exactly r because every non-zero row is linearly independent. Let $\{w_1, \dots, w_l\}$ be a basis for the kernel of Σ (the exact dimension can be computed but it is not of interest here), we claim that $\{Vw_1, \dots, Vw_l\}$ is a basis for the kernel of X. In fact, $\{Vw_1, \dots, Vw_l\}$ is a linearly independent set because V is invertible and we can write

$$X(Vw_i) = U\Sigma V^T V w_i = U(\Sigma w_i) = 0.$$

Moreover, for z in the kernel of X,

$$U^T X z = U^T U \Sigma V^T z = \Sigma V^T z = 0,$$

so $V^T z$ is a linear combination of w_i and then z can be written as a linear combination of $\{Vw_1, \dots, Vw_l\}$. We can conclude that the kernel of X has the same dimension of the kernel of Σ. By the rank-nullity theorem, X must have rank r.

- Let $X \in \mathbb{R}^{n\times m}$, assume that $n \geq m$ (the other case is analogous). By the SVD decomposition we can write

$$X^T X = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T \Sigma V^T.$$

Observe that $D := \Sigma^T\Sigma$ is a $m \times m$ diagonal matrix whose diagonal elements are $\sigma_1^2, \dots, \sigma_m^2$ and VDV^T is a spectral decomposition for the symmetric matrix $X^T X$, therefore the eigenvalues of $X^T X$ are $\sigma_1^2, \dots, \sigma_m^2$.

An alternative approach for the solution of Problem 1 is to prove the second item first and then use the fact that the number of non-zero eigenvalues is the rank of $X^T X$, since $X^T X$ has the same kernel of X, the first item follows from rank-nullity theorem.

Problem 2

Let $A \in \mathbb{R}^{n\times n}$ be a matrix with non-zero singular values ordered as usual $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$.

(a) Prove that A is invertible.

(b) Compute the condition number of the matrix A, namely $\|A\|\|A^{-1}\|$, in terms of singular values of A.

(c) Conclude that the conditional number of A is one, if and only if A is a scaled isometry.

As always, $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n | \ \|x\|_2 = 1\}$ is the standard Euclidean sphere and $\|A\| := \sup_{x\in\mathbb{S}^{n-1}} \|Ax\|_2$ is the operator norm of A.

Solution 2

(a) Observe that A is a $n \times n$, so A is invertible if and only if it has rank equal to n. By Proposition 3.1, the rank of A is the number of non-zero singular values and by assumption all singular values are non-zero.

(b) Let us first compute $\|A\|$,

$$\sup_{x\in\mathbb{S}^{n-1}} \|Ax\|_2 = \sup_{x\in\mathbb{S}^{n-1}} \sqrt{\langle Ax, Ax\rangle} = \sqrt{\sup_{x\in\mathbb{S}^{n-1}} \langle x, A^T Ax\rangle} = \sigma_1(A).$$

The last step follows from two facts: In the previous homework we saw that the maximum of the quadratic form of the right hand side is equal to the maximum eigenvalue of $A^T A$ and by Proposition 3.1 we know that it is equal to $\sigma_1(A)^2$.

Notice that the singular values of A^{-1} are $\frac{1}{\sigma_n}, \ldots, \frac{1}{\sigma_1}$. It follows that $\|A^{-1}\| = \frac{1}{\sigma_n(A)}$ and the condition number is ratio $\frac{\sigma_1}{\sigma_n}$.

(c) By the previous item, the condition number of A is one if and only if $\sigma_1 = \sigma_n = \sigma$ for some $\sigma > 0$. Moreover, $\langle Ax, Ax \rangle = \langle x, A^T Ax \rangle$ is equal to $\langle x, x \rangle$ if and only if $A^T A = I_n$, where I_n is the $n \times n$ identity matrix. If $\sigma_1 = \sigma_n = \sigma$, then by Proposition 3.1, $A^T A$ has all eigenvalues equal to σ^2 and together with the standard Spectral Theorem $A^T A = \sigma^2 I_n$, it follows that A is a scaled isometry because $\frac{1}{\sigma} A$ is an isometry by the previous discussion. The converse follows the same steps, if $A^T A$ is a multiple of the identity, then all eigenvalues are equal. By Proposition 3.1, all singular values of A are equal and then the condition number is one.

Problem 3

Let $A, B \in \mathbb{R}^{m \times n}$ be two arbitrary matrices. Find the solution, in terms of A and B, of the following optimization problem:

$$\arg \min_{\Omega \in \mathcal{O}(m)} \quad \|\Omega A - B\|_F$$

Here $\mathcal{O}(m)$ denotes the set of all $m \times m$ orthogonal matrices and $\|.\|_F$ the usual Frobenius norm.
Hint: Recall that the Frobenius norm is induced by the trace inner product in the appropriate space of matrices.

Solution 3

Let $L := \arg\min_{\Omega\in \mathcal{O}(m)} \|\Omega A - B\|_F$. We write

$$\begin{aligned} L &= \arg\min_{\Omega\in \mathcal{O}(m)} \|\Omega A - B\|_F^2 \\ &= \arg\min_{\Omega\in \mathcal{O}(m)} \langle \Omega A - B, \Omega A - B\rangle \\ &= \arg\min_{\Omega\in \mathcal{O}(m)} \|\Omega A\|_F^2 + \|B\|_F^2 - 2\langle \Omega A, B\rangle \\ &= \arg\max_{\Omega\in \mathcal{O}(m)} \langle \Omega A, B\rangle \\ &= \arg\max_{\Omega\in \mathcal{O}(m)} \langle \Omega, BA^T\rangle \end{aligned}$$

Now we consider the SVD decomposition of $BA^T = U\Sigma V^T$. It follows that

$$\begin{aligned} L &= \arg\max_{\Omega\in \mathcal{O}(m)} \langle \Omega, U\Sigma V^T\rangle \\ &= \arg\max_{\Omega\in \mathcal{O}(m)} \langle U^T\Omega V, \Sigma\rangle \end{aligned}$$

Let $W := U^T\Omega V$, by assumption and SVD decomposition we know that Ω, V and U^T are orthogonal matrices and then W is also orthogonal. The inner product $\langle W, \Sigma\rangle$ is equal to the trace of the matrix ΣW. Since Σ is a diagonal matrix with diagonal entries $\sigma_1, \dots, \sigma_m$, the inner product is equal to $\sum_{i=1}^m \sigma_i W_{ii}$. By the previous discussion, W is an orthogonal matrix and then $W_{ii} \leq 1$. It follows that $\langle W, \Sigma\rangle \leq \sum_{i=1}^m \sigma_i$ and the equality holds if and only if W is $m \times m$ the identity matrix. Finally, for $\Omega = UV^T$, we have $W := I_m$, so $L = UV^T$.