Homework 2

Ercoli, Fabio Massimo
802397

Mai Mihai, Cristian Andrei
1942925

December 21st 2024

1 Assignment 1

1.1 SVD

For every matrix $A \in \mathbb{R}^{m\times n}$, having rank r, there always exist its *Singular Value Decomposition*:

$$A = U\,\Sigma\,V^T \tag{1}$$

Where:

- The left singular vectors are the columns of the matrix U. They form an orthonormal basis in $\mathbb{R}^n$.
- The right singular vectors are the columns of the matrix V. They form an orthonormal basis in $\mathbb{R}^m$.
- $\Sigma = diag(\sigma_1, \sigma_2, ..., \sigma_{\mathrm{r}})$, where $\{\sigma_{\mathrm{i}}\}$ are the singular values.

1.2 Eigenvectors of AA^T and A^TA

For any rectangular matrix A, we can always say that:

- the left singular vectors are the eigenvectors of the matrix AA^T associated to the all non zero eigenvalues of AA^T.
- the right singular vectors are the eigenvectors of the matrix A^TA associated to the all non zero eigenvalues of A^TA .

We used those values as columns of U and V to prove the existence and uniqueness of the *SVD* of A.

Alternatively, those claims can be also derived starting from the existence of the *SVD* of A. That is what we're going to do now.

Notice that we can write A^T as:

$$A^T = ((U\,\Sigma)\,V^T)^T = V(U\,\Sigma)^T = V\Sigma^T U^T = V\Sigma U^T \tag{2}$$

The last equality stands because Σ is diagonal and then symmetric.

$$AA^T = U\ \Sigma\ V^T\ V\Sigma\ U^T = U\ \Sigma^2\ U^T \tag{3}$$

$$A^TA = V\ \Sigma\ U^T\ U\Sigma\ V^T = V\ \Sigma^2\ V^T \tag{4}$$

Multiplying both sides of the equation (3) by U and the (4) by V, we get:

$$AA^T \cdot U = \Sigma^2\ U^T \Longrightarrow AA^T\mathbf{u_i} = \sigma^2 \cdot \mathbf{u_i}, \forall i \in 1..r \tag{5}$$

$$A^TA \cdot U = \Sigma^2\ V^T \Longrightarrow A^TA\mathbf{v_i} = \sigma^2 \cdot \mathbf{v_i}, \forall i \in 1..r \tag{6}$$

That proves the claims.

1.3 Solution to the point 1(b)

Since A is symmetric, we know that:

$$AA^T = A^TA = A^2 \tag{7}$$

Using the result of the last chapter we can say that:

- the left singular vectors are the eigenvectors of the matrix A^2 associated to the all non zero eigenvalues of A^2.
- the right singular vectors are the eigenvectors of the matrix A^2 associated to the all non zero eigenvalues of A^2 .

That implied that left and right singular vectors are the same.

1.4 Eigenvectors of A

Let's denote $\boldsymbol{x_i}$ the i-th eigenvector of A. By definition, we know that:

$$\exists \lambda_{\mathrm{i}} : A \cdot \mathbf{x_i} = \lambda_{\mathrm{i}} \cdot \mathbf{x_i} \tag{8}$$

Multiplying both sides by A we have:

$$A^2 \cdot \mathbf{x_i} = \lambda_{\mathrm{i}} \cdot A \cdot \mathbf{x_i} \tag{9}$$

Applying the (8) again to the right size of (9) we have:

$$A^2 \cdot \mathbf{x_i} = {\lambda_{\mathrm{i}}}^2 \cdot \mathbf{x_i} \tag{10}$$

So the eigenvectors $\boldsymbol{x_i}$ are the same as singular vectors, in case of course the matrix A is symmetric.

1.5 Singular values and eigenvalues

The i-th singular value σ_i of the matrix A can be computed as the square root of the corresponding eigenvalue of AA^T, in our case also of the corresponding eigenvalue of A^2.

Looking at the (10), we can notice that i-th eigenvalue of A^2 is λ^2, so the corresponding (i-th) singular value is λ, that is the (i-th) eigenvalue of A.

So we can conclude that also eigenvalue and singular values of A are the same.

2 Assignment 2

To calculate *E[R_{ij}]* we can apply the definition of expected value of a random variable.

$$E[R_{\text{ij}}] = \sum_x x \cdot P[R_{\text{ij}} = x] \tag{11}$$

where x varies over all possible values in the range of x. According to the definition of R_{ij}:

$$E[R_{\text{ij}}] = -\sqrt{\frac{s}{k}}\frac{1}{2s} + 0(1-\frac{1}{s}) + \sqrt{\frac{s}{k}}\frac{1}{2s} = 0 \tag{12}$$

To say something about $E[\|f(u)\|^2]$, we need first to compute $E[{R_{\text{ij}}}^2]$, we will reuse the result later. Since:

$${R_{\text{ij}}}^2 = \begin{cases} \frac{s}{k}, & \text{with prob.}\frac{1}{s}; \\ 0, & \text{with prob.}1-\frac{1}{s}; \end{cases} \tag{13}$$

We can say that:

$$E[{R_{\text{ij}}}^2] = \frac{s}{k}\frac{1}{s} + 0(1-\frac{1}{s}) = \frac{1}{k} \tag{14}$$

$$E[\|f(u)\|^2] = E[\|R^T\mathbf{u}\|^2] = \sum_{i=1}^{k} E[(R(i,*)\mathbf{u})^2] \tag{15}$$

where $R(i,*)$ it the i-th row of the matrix R^T. *R(i,*)* has dimension (1 x d), while $\mathbf{u}$ has dimension (d x 1), so the result of their product $R(i,*)\mathbf{u}$ is a scalar value. Since the random variable R_{ij} does not depend on the row number, we can say that:

$$E[\|f(u)\|^2] = k \cdot E[(\mathbf{r}^T \cdot \mathbf{u})^2] \tag{16}$$

Where we have replaced *R(i,*)* with $\mathbf{r}^T$, to indicate a generic vector of size d having as values the R_{ij}.

$$
\begin{aligned}
E[(\mathbf{r}^T \cdot \mathbf{u})^2] = E\left[\left(\sum_{i=1}^{d} r_{\mathrm{i}}\ u_{\mathrm{i}}\right)^2\right] \\
= E\left[\sum_{i=1}^{d}(r_{\mathrm{i}}\ u_{\mathrm{i}})^2 + \sum_{l=1}^{d}\sum_{m=1}^{d} 2r_{\mathrm{l}}r_{\mathrm{m}}u_{\mathrm{l}}u_{\mathrm{m}}\right] \\
= \sum_{i=1}^{d} E[{r_{\mathrm{i}}}^2]{u_{\mathrm{i}}}^2 + \sum_{l=1}^{d}\sum_{m=1}^{d} 2E[r_{\mathrm{l}}]E[r_{\mathrm{m}}]u_{\mathrm{l}}u_{\mathrm{m}} \\
= \frac{1}{k} \cdot \|u\|^2
\end{aligned}
\tag{17}
$$

In which we applied that $E[{r_{\mathrm{i}}}^2] = \frac{1}{k}$, $\sum_{i=1}^{d} {u_{\mathrm{i}}}^2 = \|u\|^2$ and $E[r_{\mathrm{i}}] = 0$.
Finally, we can use the result and compute:

$$
E[\|f(u)\|^2] = k\frac{1}{k}\|u\|^2 = \|u\|^2 \tag{18}
$$

3 Assignment 3

After we've processed n pairs of the stream $[(\mathrm{a}_1, \mathrm{w}_1) \ldots (\mathrm{a}_{\mathrm{n}}, \mathrm{w}_{\mathrm{n}})]$, we want that the probability of any pair $(\mathrm{a}_{\mathrm{i}}, \mathrm{w}_{\mathrm{i}})$, with $\mathrm{i} \leq \mathrm{n}$, is in the sample is exactly equal to $\frac{w_{\mathrm{i}}}{W_{\mathrm{n}}}$, where $\mathrm{W}_{\mathrm{n}} = \sum_{j=1}^{n} w_{\mathrm{j}}$.

3.1 Define the algorithm

The algorithm can be inductively defined. The base is case is when the first pair of the stream $(\mathrm{a}_1, \mathrm{w}_1)$ arrives, in this case we initialize the variables:

$$
sample = a_1; W = w_1 \tag{19}
$$

The other case is when we already seen n pairs of the stream, with n grater than 0, and the new pair $(\mathrm{a}_{\mathrm{n+1}}, \mathrm{w}_{\mathrm{n+1}})$ pair arrives. The central idea is to choose as sample the incoming $(\mathrm{a}_{\mathrm{n+1}}, \mathrm{w}_{\mathrm{n+1}})$ pair with probability exactly equals to $\frac{w_{\mathrm{n+1}}}{W + w_{\mathrm{n+1}}}$.

If we choose it, we update the variable *sample* to $a_{\mathrm{n+1}}$. Otherwise, we keep the *sample* as it is. Either way we update the variable W to $W + w_{\mathrm{n+1}}$.

3.2 Prove that the algorithm is correct

We can prove by induction on n.

In the base case, n=1, the probability that a_1 is the sample variable is 1. Then it is equal to w_1/W_1, so we achieves the desired probabilistic guarantees stated above for the only element corresponding to $i = 1$.

Now let's consider the inductive step in which we process the *(n+1)*-st pair of the stream. The probability that the *(n+1)*-st pair is selected is equal to:

$$\frac{w_{\mathrm{n+1}}}{W_{\mathrm{n}} + w_{\mathrm{n+1}}} = \frac{w_{\mathrm{n+1}}}{W_{\mathrm{n+1}}} \tag{20}$$

Then the algorithm has the desired probabilistic guarantees also in this case.

About the other pairs in the first n positions, applying the induction hypothesis, we know that they all have probability equals to $\frac{w_{\mathrm{i}}}{W_{\mathrm{n}}}$ to be in the sample just before the *(n+1)*-st pair is processed.

If the *(n+1)*-st pair is chosen, the probability that the i-th pair, with $1 \leq i \leq n$, is in the sample is simply zero. In this setting (s = 1).

While the probability of having the i-th pair at the end of the step if the *(n+1)*-st pair is **not** chosen is equals to $\frac{w_{\mathrm{i}}}{W_{\mathrm{n}}}$.

To compute the probability of having the i-th pair in the sample at the end of the step (e.g., just after the *(n+1)*-st pair is processed) we can use the law of total probability conditioning the result on the fact that *(n+1)*-st pair may be chosen or not.

Putting all tougher we have:

$$\begin{aligned} &\left(1 - \frac{w_{\mathrm{n+1}}}{W_{\mathrm{n}} + w_{\mathrm{n+1}}}\right)\left(\frac{w_{\mathrm{i}}}{W_{\mathrm{n}}}\right) + \left(\frac{w_{\mathrm{n+1}}}{W_{\mathrm{n}} + w_{\mathrm{n+1}}}\right) \cdot 0 \\ &= \frac{W_{\mathrm{n}} + w_{\mathrm{n+1}} - w_{\mathrm{n+1}}}{W_{\mathrm{n}} + w_{\mathrm{n+1}}}\left(\frac{w_{\mathrm{i}}}{W_{\mathrm{n}}}\right) = \frac{w_{\mathrm{i}}}{W_{\mathrm{n}} + w_{\mathrm{n+1}}} = \frac{w_{\mathrm{i}}}{W_{\mathrm{n+1}}} \end{aligned} \tag{21}$$

That should complete the proof.