Linear Data Analysis
Matricies and Linear Regression Matrix Approximation

Cain Susko

Queen's University
School of Computing

February 28, 2022

a Matrix Norms

this section will focus on how to use one matrix to approximate another matrix. We will look at the Axioms of a matrix.

Matrix Norms: Axioms how close are 2 matricies ? what is $||A - C||$? consider $A \in R^{m \times m}, C \in R^{m \times m}, a \in R$. the matrix norm $|| \cdot ||$ satisfies 4 axioms:

- $||A|| \geq 0$
- $||A|| = 0$ iff $A = 0$
- $||\alpha A|| = |\alpha|; \ ||A||$
- $||A + C|| \leq ||A|| + ||C||$

additionally, compatible matrix and vector norms for $\vec{w} \in R^n$

$$||A\vec{w}|| \leq ||A|| \ \ ||\vec{w}||$$

b L2 Matrix Norm and Frobenius Matrix Norm

this section will explore 2 types of norms.

L_2 **Norm** For $A \in R^{m \times m}, \vec{w} \in R^n$ the L_2 norm $||A||_2$ is defined as:

$$||A||_2 =^{def} \frac{||A\vec{w}||}{||\vec{w}||}$$

such that $||A||_2$ is the largest possible value that can be computed given the equation and A. in summary, if we find the largest eigenvalue λ_{max}, then

$$||A||_2 = \sqrt{\lambda_{max}(A^\top A)} = \sigma_1$$

Frobenius Norm for $A \in R^{m \times n}$:

$$||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij})^2} = \sqrt{\sigma_1^2, \sigma_2^2 ..., \sigma_n^2} = \vec{\sigma}$$

$$||A||_F^2 = \sum_{i} \sum_{j} (a_{ij})^2 = \vec{\sigma}^\top \vec{\sigma}$$

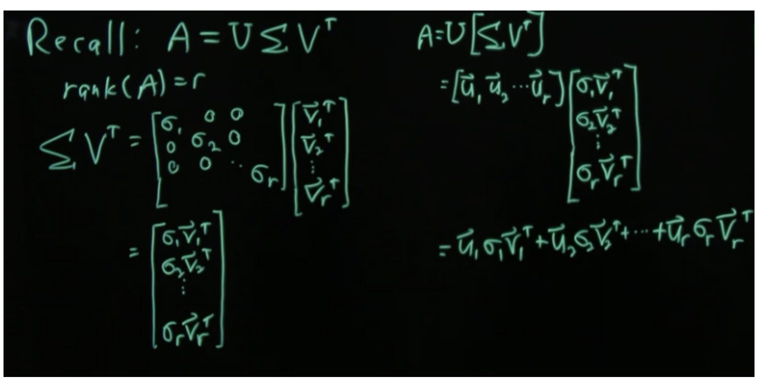
thus we can see the L_2 norm is the largest singular value (σ_1) and the Forbenius Norm is the Euclidian Norm of the vector of singular values ($\vec{\sigma}$)

c Matrix Series from the SVD

this section will focus on how to write a matrix as a series. there are some applications that require this simplification.
Recall that the SVD is:

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

where $rank(A) = r$ the matricies would look like the following:



thus we can see how the SVD is changed into a different decomposition of

$$A = U[\Sigma V^\top]$$

this thus (it does) show us that any matrix A with rank r can be summed as many rank 1 matricies using the above equation. (iff $\vec{u} \neq 0, \vec{v} \neq 0$)

d Rank-k approximations

we will cover how to approximate a matrix with a rank-k marix.
Given: $A \in R^{m\times n}$, we will approximate A with $C \in R^{m\times n}$. We first want to measure $||A - C||$. Consider $rank(C) = 1$, build C from $\vec{z} \in R^m$, thus:

$$C = [\alpha_1\vec{z}, \alpha_2\vec{z}, ..., \alpha_n\vec{z}] = \vec{z}\vec{\alpha}^\top$$

Alternatively; we can use $||\vec{w}|| = 1$

$$\vec{\alpha} = \beta\vec{w} \quad C = \vec{z}\beta\vec{w}^\top$$

but what are the optimal values of $\vec{z}, \beta, \vec{w}$ with the L_2 and Frobenius norm:

$$C = \vec{u}_1\sigma_1\vec{v}_1^\top$$

Once again, consider $A = C_1 + C_2 + ... + C_r$ where each C_i is a rank one matrix (from using the matrix series equation from section c of this note. Thus we can define C_i as

$$C_i =^{def} \vec{u}_i\sigma_i\vec{v}_i^\top$$

The Eckart-Young Theorum states that the optimal rank k approximation for A is $C = C_1 + C_2 + ...C_k$ where C_i is derived from the SVD of A.
Note that, the column space of $C_1 + C_2$ is equal to $U = [\vec{u}_1\vec{u}_2]$. This thus means that a rank-k approximation of A is *also* a rank-k approximation of the column space of A.

Scree Plot

this section explores guidelines for matrix approximation.
Given $A = U\Sigma V^\top$, form $\vec{\sigma}$.
We first want to rescale $\sigma \in [0, 1]$. to do this, we must find the explaiined variance

$$\Theta = \sum_{i=1}^{r} \sigma_1 = ||\vec{\sigma}||_1$$

and the total variance

$$T = \sum_{i=1}^{r} (\sigma_i)^2 = ||\vec{\sigma}||_i^2$$

if we plot $\vec{\sigma}/\Theta$ or $\vec{\sigma}/T$ we then get a **Scree plot** where the interesting thing to us is the *elbow* if it has one.

Examples mathematically, we can see that as we progress through i from C_i, we can see that the approximation of A gets better, and that a rank 2 approximation of A is pretty good

$$\text{Consider } A = \begin{bmatrix} 5 & 5 & 2 \\ 5 & 4 & 3 \\ 1 & 1 & 5 \end{bmatrix}$$

$$\text{Singular values are} \approx 10.68, 4.07, 0.53$$

$$C_1 \approx \begin{bmatrix} 4.70 & 4.26 & 3.36 \\ 4.62 & 4.18 & 3.30 \\ 2.35 & 2.13 & 1.68 \end{bmatrix}$$

$$C_1 + C_2 \approx \begin{bmatrix} 5.23 & 4.75 & 2.00 \\ 4.73 & 4.29 & 3.00 \\ 1.07 & 0.92 & 5.00 \end{bmatrix}$$

visually in this second example, we can see the elbow mentioned earlier: this line represents the fit for the approximation of A by C_j. The elbow shows where the approximation gets closer.