MATH-BIOINF-STATS 547: Mathematics of Data

Due Date: Thursday, January 30, 2025

Problem Set 1: Singular Value Decomposition and Data

For this problem set, please submit a `.pdf` document with a write-up of your results and observations. We encourage using Overleaf, but MS Word or other similar word-processing software is OK. We have provided a LaTeX template to help get you started, which is available on Overleaf where you can make a copy of it.

Problem 1

(a) Compute $\mathbf{A}^\mathsf{T}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\mathsf{T}$ and their eigenvalues and unit eigenvectors for $\mathbf{V}$ and $\mathbf{U}$.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Check if $\mathbf{AV} = \mathbf{U\Sigma}$ (this will decide $\pm$ signs in $\mathbf{U}$). $\mathbf{\Sigma}$ has the same shape as $\mathbf{A}$

(b) Compute the singular values and eigenvalues of the following matrix. Explain your observations.

$$\mathbf{A} = \begin{bmatrix} -149 & -50 & -154 \\ 537 & 180 & 546 \\ -27 & -9 & -25 \end{bmatrix}$$

(c) Repeat part (b) with the following matrix. Do you observe a similar relationship between the singular values and eigenvalues? What could explain any changes to this relationship?

$$\mathbf{A} = \begin{bmatrix} -18 & 150 & -40 \\ 150 & 200 & 60 \\ -40 & 60 & -116 \end{bmatrix}$$

Problem 2

In this problem, you will explore the 'MATH: 547 Introductions' class data generated from your responses to the introduction survey. You will generate visualizations, compute the singular value decomposition (SVD) on real data and document your observations. We encourage you to be creative and have fun when plotting and presenting your results. Let the data matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$ with m students as the rows and n features as the columns. You may use the helpful code provided as a template.

(a) Load the file '`MATH 547 Introductions.csv`' into MATLAB (or another programming language). How many rows and columns does the data have? What is the rank of $\mathbf{A}$?

(b) On average, are students more excited by 'singular value decomposition' or by 'machine learning'? On average, are students more excited by 'working with real data' or with programming?

(c) Create a plot showing the relationship between students' interests and their current skills. You may choose any relationship you like, and any plot type you like. Please explain your observations from the plot you generate.

(d) Compute SVD of $\mathbf{A} = \mathbf{U\Sigma V}^T$. What are the sizes of the matrices $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}$?

(e) Create a scree plot of the singular values of $\mathbf{A}$. Does this change your mind about the rank of $\mathbf{A}$ from part (a)?

(f) Compute $\mathbf{A}^\mathsf{T}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\mathsf{T}$. Plot both $\mathbf{A}^\mathsf{T}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\mathsf{T}$. One option is to use MATLAB's function `imagesc`. Please explain your observations. What is the size of each of these matrices? What do these matrices represent?

(g) Take the SVD of $\mathbf{A}^\mathsf{T}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\mathsf{T}$. Compare the values in $\mathbf{\Sigma}$ from each matrix. What do you notice? How do these values relate to the singular values of $\mathbf{A}$ from part (d)?