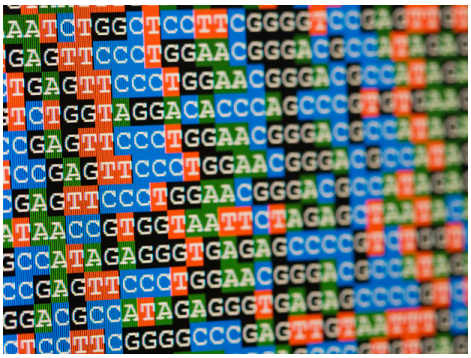
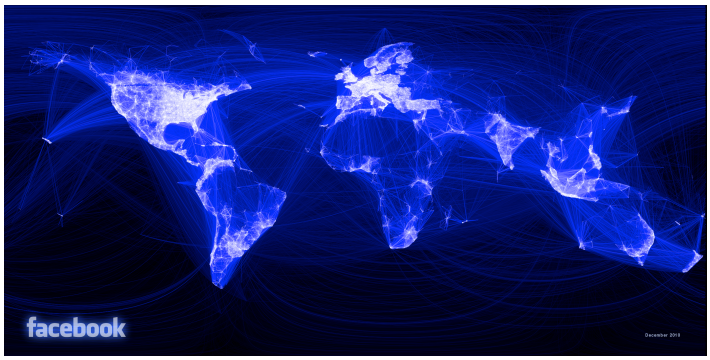
MATH 496T: Spring 2020

Advanced Linear Algebra for Data Science

Instructor: Dr. Keaton Hamm:

hamm@math.arizona.edu





About the Course: We are collecting a phenomenal amount of data these days from every endeavor that human beings are involved in. But harnessing this data requires sophisticated techniques due to its large scale and complexity. At a fundamental level, most data is represented as vectors in Euclidean space $\mathbb{R}^n$ or $\mathbb{C}^n$ where n may be on the order of millions. Thus, Linear Algebra provides some of the core tools for Data Analysis.

This course will cover some of the more advanced topics in Linear Algebra beyond what is covered in traditional undergraduate courses. The focus will be on exploring theory which is used in real Data Science applications, including matrix factorization, low-rank matrix approximations, kernel methods, graph theory, and optimization. Some special modern topics will be covered such as Compressed Sensing, Data Clustering, and Frame Theory. The theory will be complemented with illustrative applications.

Prerequisites: This course will be accessible to anyone who has taken MATH 313 or 310 (Introduction to Linear Algebra, or Applied Linear Algebra, respectively). Some programming experience is beneficial as there will be both written homeworks as well as programming assignments (students will be allowed to use any coding language they like). I will work with students with little coding experience to make sure the assignments are manageable. There will be some overlap with the Numerical Analysis sequence as we will discuss the computational issues arising in analyzing large-scale, high-dimensional data.

Textbook: *Linear Algebra and Learning from Data* by Gilbert Strang, `https://math.mit.edu/~gs/learningfromdata/`

Objectives: Students will learn

- How to compute various matrix factorizations and understand their quantitative and qualitative differences and similarities
- How to use optimization methods for solving linear systems of equations and for reducing dimensionality of data sets
- How to use Kernel matrices to cluster data
- What applications utilize the methods discussed.