STA 141C - Big Data & High Performance Statistical Computing

Bo Y.-C. Ning

2022 Winter Quarter @ UC Davis

January 4, 2022


2022
Happy new year and welcome to STA141C !

Introduction

- This course is **not** about teaching you how to write R/python programming code
- This course is **not** about teaching you how to use packages to fit statistical models

What is this course about?

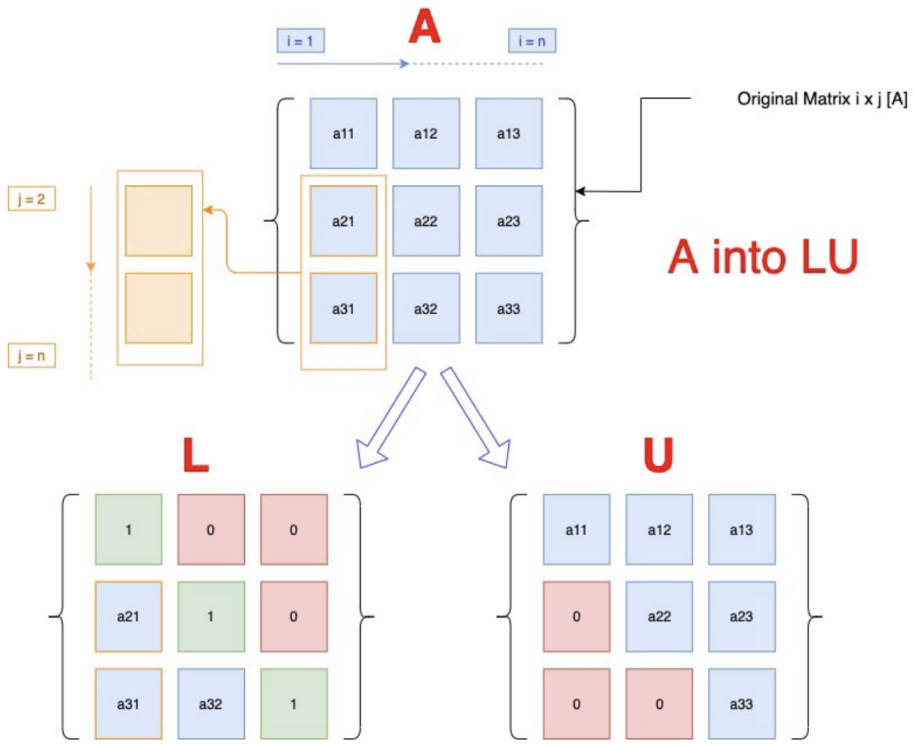
1. Numeric linear algebra methods

Statistical methods often require to solve

$$A\beta = b$$

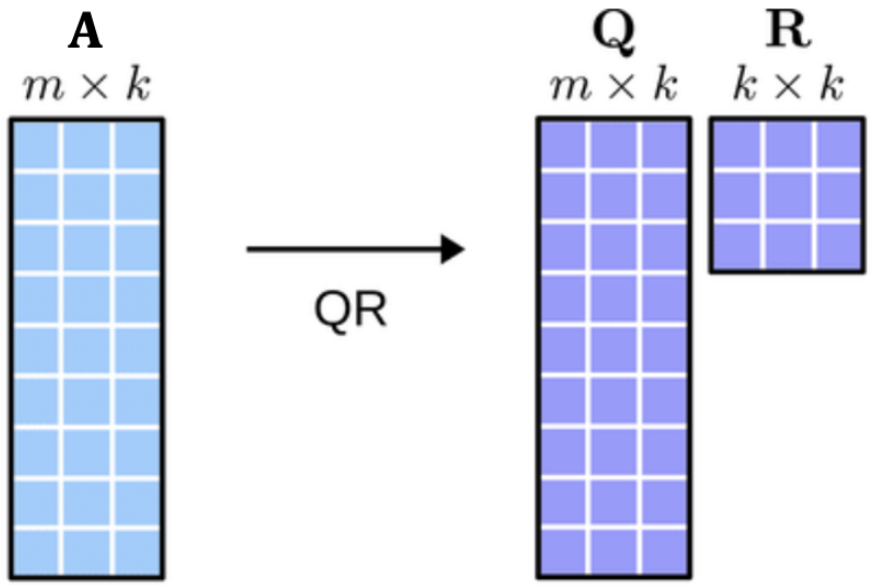
We will learn different methods to solve β efficiently

Example 1: LU decomposition

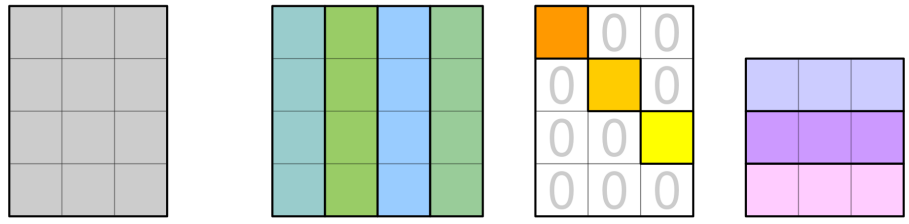


source: `https://towardsdatascience.com/`

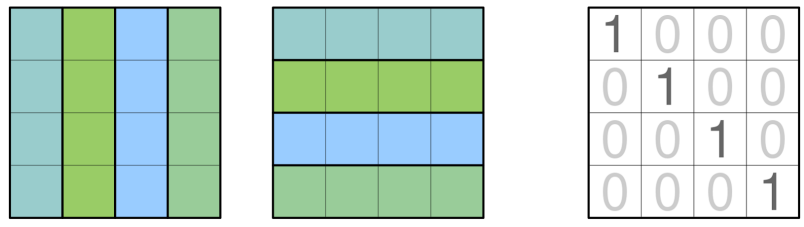
Example 2: QR factorization (decomposition)



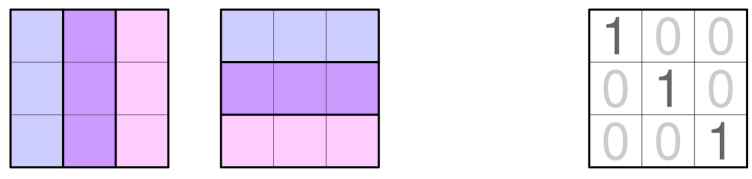
Example 3: SVD (Singular value decomposition)



$$\underset{m\times n}{\mathbf{M}} = \underset{m\times m}{\mathbf{U}}\ \underset{m\times n}{\mathbf{\Sigma}}\ \underset{n\times n}{\mathbf{V}^*}$$



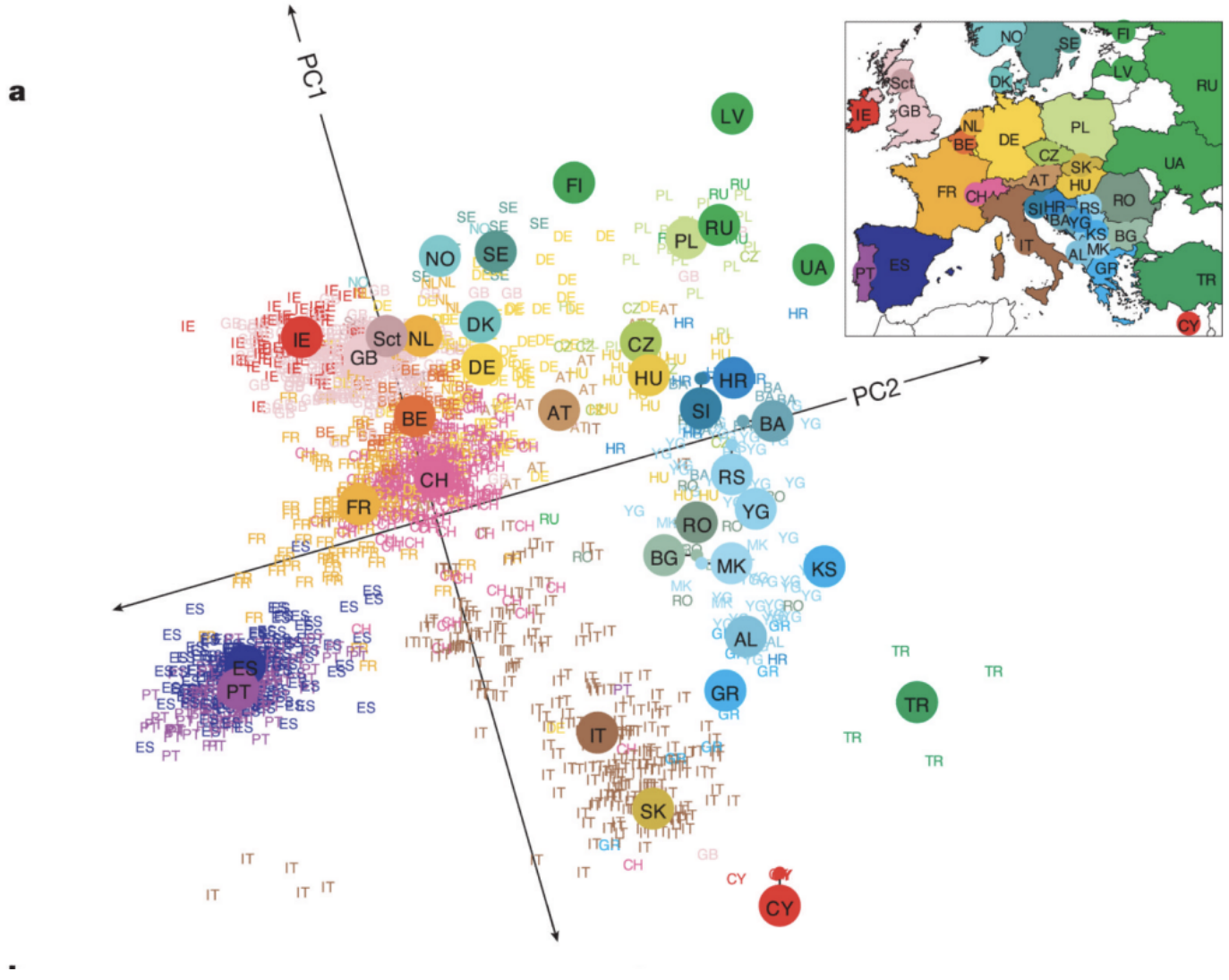
$$\mathbf{U}\ \mathbf{U}^* = \mathbf{I}_m$$



$$\mathbf{V}\ \mathbf{V}^* = \mathbf{I}_n$$

Source: wikipedia

Application: PCA (Principal component analysis)

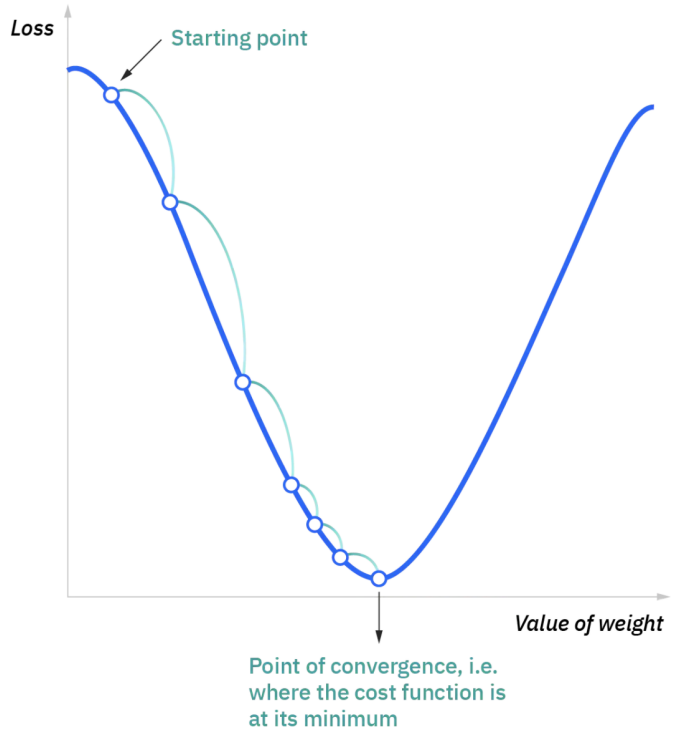


Source: Genes mirror geography within Europe. *Nature*. 2008

- So, you need to know basic linear algebra (of course, I will review many of them, but it is better that you already knew them)
- You also need to know how to write code (I will not teach you how to write code). For completing homework assignments, you are free to choose R and python or both

2. Algorithms for statistical models for big data

Example 1: Optimization for nonlinear functions



source: `https://www.ibm.com/cloud/learn/gradient-descent`

A list of algorithms: Newton's method, Gradient descent, etc

Application: Neural Networks

When Neural Networks saw the first image of Black Hole.

Anuj shah (Exploring Neurons) Follow

Apr 19, 2019 · 6 min read

Black Hole — M87 — Event Horizon Telescope

source: `https://medium.com/analytics-vidhya/when-neural-networks-saw-the-first-image-of-black-hole-3205e28b6578`

Example 2: Sampling based methods for Bayesian methods

Example 2: Sampling based methods for Bayesian methods

Application: Gaussian Processes

- Transit method for finding exoplanets: `https://www.youtube.com/watch?v=xNeRqbw18Jk`
- Stellar activity: `https://www.physics.uu.se/research/astronomy-and-space-physics/research/stars/magnetism/`

Example 2: Sampling based methods for Bayesian methods

Application: Gaussian Processes

- Transit method for finding exoplanets: `https://www.youtube.com/watch?v=xNeRqbw18Jk`
- Stellar activity: `https://www.physics.uu.se/research/astronomy-and-space-physics/research/stars/magnetism/`

IMPROVING EXOPLANET DETECTION POWER: MULTIVARIATE GAUSSIAN PROCESS MODELS FOR STELLAR ACTIVITY

By David E. Jones[1], David C. Stenning[2] Eric B. Ford[3], Robert L. Wolpert[4], Thomas J. Loredo[5], Christian Gilbertson[3], and Xavier Dumusque[6]

Texas A&M University[1], Simon Fraser University[2], Penn State University[3], Duke University[4], Cornell Center for Astrophysics and Planetary Science[5], and Observatoire Astronomique de l'Université de Genève[6]

You will learn ...

- Newton's method
- Gradient decsent
- EM algorithm
- Variational inference
- Bayesian methods and Markov chain Monte Carlo
- Parallel computing
- ....

Annoucement

- My office hour this is 1:00-2:00pm this afternoon
- No TA office hour this week
- Install Rstudio or the Jupyter Notebook
- If you have any issue, Wei can help you during her discussion session tomorrow