Reading list

Session 5

Magnus Nielsen

Please note that you are **not** expected to have read the reading materials before the classes and should be able to follow along without having done so.

Raschka, S., & Mirjalili, V. (2019). Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing Ltd.:

- Chapter 5, p. 145-159, Compressing Data via Dimensionality reduction, section about Principal Component Analysis

    – Focus on first two pages, rest are cursory reading or if interested in the math

- Chapter 11, p. 353-358, Working with Unlabeled Data - Clustering Analysis, section about K-means, until k-means++
- Chapter 11, p. 361-362, Working with Unlabeled Data - Clustering Analysis, Using the elbow methods to find the optimal number of clusters
- Chapter 11, p. 363-367, Working with Unlabeled Data - Clustering Analysis, Quantifying the quality of clustering via silhouette plots
- Chapter 12, p. 262-271, Applying Machine Learning to Sentiment Analysis, Introducing the bag-of-words model

    – Cursory reading

For those that are interested in text as data, I can also recommend Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. Journal of Economic Literature, 57(3), 535-74.. It includes a lot of inspiration