MTH6101: Introduction to Machine Learning Semester B, 2023-24

Essentials for 1st quiz

Kostas Papafitsoros & Hugo Maruri-Aguilar

When?

Due to multiple clashes (Week 7 is partially teaching week, activities arranged by students...), we come back to the original day:

Friday Week 7, March 8: 16:00-18:00

- No labs that day!

Structure

- Purely online assessment in a form of a QMplus quiz
- PCA and clustering
- Three questions (randomised): every students will have their own unique data

Help during preparation

- A pdf with sample questions and solutions will be provided this week
- Teams channel: "MTH6101 Lecture chat"
  Click **here** for a link to the channel
  Password: akzwxmy
  Questions slot: Wednesday March 6, 15:00-17:00

Essential knowledge

Here is a list of essential tasks you should be ready to do:

- Given a dataset X, you should be able to compute percentages of total variability explained by a given principle component (e.g. PC1, PC2 etc).

- Given a dataset X, and a threshold of total variability, you should be able to decide how many principle components you should choose.

- Given a dataset X, and its variance matrix you should be able to determine whether this dataset was scaled or not.

- Given a biplot, you should be able to interpret the first and second principal component (weighed averages, comparisons, large/small contributions of variables).

- Given a biplot, you should be able to make suggestions on whether the data were scaled or not.

- Given some data X, you should be able to compute the distance matrix under a specific metric.

Essential knowledge

- Given some data X and some specific clusters you should be able to compute the dissimilarity (distance) between these clusters under "single"/average/complete" linkage.

- You should be able to compute the centroid of a cluster (mean or medoid).

- You should be able to understand, describe and interpret a dendrogram. In particular you should be able to suggest the most appropriate clusters based on the dendrogram heights.

- You should be able to interpret silhouette plots and suggest most appropriate number of clusters based on these plots.

Also look at the boxes at the end of every chapter in the notes:

> Important points and concepts of week ...
>
> 1. ...
>
> 2. ...