

Лабораториска Вежба 1

Класификација со Naïve Bayes и Gaussian Discriminant Analysis модели (LDA и QDA)

1. Изберете податочно множество со кое ќе работите

- a. <https://archive.ics.uci.edu/ml/datasets/Abalone> - повеќекласна класификација на старост на Abalone (вид на гастропод).
- b. <https://archive.ics.uci.edu/ml/datasets/Glass+Identification> – повеќекласна класификација на вид на стакло.
- c. <https://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope> – бинарна класификација на гама честички во однос на шум.
- d. <https://www.kaggle.com/datasets/wenruiiu/adult-income-dataset> - бинарна класификација на приход на секој испитаник.
- e. <https://www.kaggle.com/datasets/amirhosseinzinati/wine-recognition-data/data> - повеќекласна класификација на вид на вино.
- f. <https://www.kaggle.com/datasets/henriqueyamahata/bank-marketing/data> - бинарна класификација дали е поднесен депозит од страна на клиент.

2. Запознавање со податочното множество.

Направете соодветни визуелизации со кои би добиле претстава за атрибутите на дадена класа и за тоа како меѓусебно се поврзани класите. Отстранете ги класите кои сметате дека не се потребни за вашето предвидување доколку според вас дека има такви.

3. Класификација

Поделете го оригиналното множество на тренинг и тест и соодветно истренирајте три класификатори (Naïve Bayes, LDA и QDA) користејќи ја библиотеката Scikit-Learn.

4. Споредба на моделите за класификација

Споредете ја точноста (ассигасу) помеѓу различните модели за класификација. Пробајте да заклучите зошто еден метод е подобар од друг во однос на претпоставките кои ги користат методите.

На курсот задолжително поставете го .ipynb фајлот со решение.