

Лабораториска Вежба 2

Линеарна Регресија, Lasso Регресија и Ridge Регресија

1. Изберете податочно множество со кое ќе работите.

- <https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression/data>
- <https://www.kaggle.com/datasets/yasserh/song-popularity-dataset>
- <https://www.kaggle.com/datasets/swatikhedekar/price-prediction-of-diamond/data>
- <https://www.kaggle.com/datasets/shibumohapatra/house-price>
- <https://www.kaggle.com/datasets/brsahan/vehicle-co2-emissions-dataset/data>

2. Запознавање со податочното множество и претпроцесирање.

Направете соодветни визуелизации со кои би добиле претстава за атрибутите во однос на таргет променливата. За таа цел можете да ја користите ф-јата `jointplot` од `seaborn` библиотеката, како и други визуелизации достапни во рамките на `seaborn` и `matplotlib` библиотеките. Дополнително направете `min-max` нормализација на податоците.

3. Регресија.

Истренирајте ги следните регресиони модели:

- `LinearRegression`
- `Ridge`
- `Lasso`
- `ElasticNet`
- `BayesianRidge`

Поделете го оригиналното податочно множество на тренинг и тест множество и истренирајте го секој од регресионите модел на тренинг множеството. Притоа за разлика од примерите од часовите каде што се користат стандардните `Lasso` и `Ridge` класификатори од `Scikit-Learn`, во оваа лабораториска вежба треба да ги искористите `LassoCV` и `RidgeCV` класификаторите. Целта на овие класи е тоа што во нив е веќе изграден метод на `cross-validation` кој на самото тренинг множество го оптимизира хипер-параметарот `alpha` на овие модели кој служи за ниво на регуларизација во двата модели. Со цел сами да не го избираме овој хипер-параметар, со овие класи можеме да го најдеме тој што ќе има најсоодветна вредност во однос на податоците кои ги тренираме. Откако ќе ги истренирате регресорите, пресметајте и споредете го `MSE` (`Mean Square Error`) на тест множеството за секој од нив, приметете дали има некоја разлика во коефициентите на моделите, како и која е соодветната вредност за хипер-параметарот на регуларизација во `Lasso` и `Ridge`.

На курсот задолжително поставете го `.ipynb` фајлот со решение.