

Увод у вештачку интелигенцију

Поправни колоквијум – машинско учење

21. 1. 2024.

Потребно је решити класификациони проблем *Breast Cancer*. Учитати сет података из библиотеке *sckit-learn*. Проверити да ли су сви подаци исправни. Уколико има неисправних података одбацити те редове. Податке делити на део за обуку и део за тест. Улазне податке нормализовати. Из података за обуку издвојити тренинг и валидациони скуп. Имплементирати класу *LogistickaRegresija*. UML дијаграм класе дат је у наставку.

C LogistickaRegresija	
lr: float iterations: int histoty: Any	
__init__(self, lr: float, iterations: int) -> None fit(self, X_train : np.ndarray, y_train: np.ndarray, X_valid: np.ndarray, y_valid: np.ndarray) -> 'LogistickaRegresija': predic(self, X : np.ndarray) -> np.ndarray: predict_proba(self, X: np.ndarray) -> np.ndarray:	

Метода за обуку прихвата и валидациони скуп. Оценити модел на тренинг и валидационом скупу у току обуке после сваке итерације метриком *Log loss*. Одређивање тежинских коефицијената извршити алгоритмом обуке перцептрона. Метода *predict_proba* даје „вероватноће“ у предикцији позитивне класе, док метода *predict* даје вредности класа.

Обучити креирани модел. Нацртати криве учења на тренинг и валидационом скупу. На основи графика дати коментар о могућим проблемима *underfitting* и *overfitting*.

На тестном скупу оценити модел метриком *Recall*.

Креирати модел типа *sklearn.linear_model.LogisticRegression* над истим тренинг подацима. Упоредити моделе на тестном скупу на основу ROC кривих на истом дијаграму.

Додатак:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$$