Due date: Wednesday, November 23, 2022, 12:00

1 Predict housing prices

Use the California housing dataset to predict housing prices. (https://scikit-learn.org/stable/datasets/real_world.html#california-housing-dataset)

1.1 Task 1 – Linear Regression (7 pt)

Which feature correlates most with the medium house value according to a linear model? Motivate your answer by estimating the average R2 score of each model. How do you estimate the average R2 score of a model?

1.2 Task 2 – MLP Regression (7 pt)

What is the best R2 score obtained with an MLP given the set of provided hyperparameters? Report the corresponding parity plot evaluated on the test set, the best hyperparameters, describe how you optimized them, and motivate your choices. Use the parity plot to discuss the performances of the best model. Consider standardizing the input features. Why is it important when training an MLP?

2 Predict tumor state (6 pt)

Train an MLP using the provided hyperparameters to predict the type of observed tumor using all the available features and show the ROC curve (https://en.wikipedia.org/wiki/Receiver_operating_characteristic). What is the accuracy of your model ? Train a new model only using the features 'mean area' and 'mean concavity', and compare the ROC curves: what are a false and true positive in the context of a diagnosis? Which model is best suited for this task and why?