Exercise   Regression algorithms

Ames Housing Dataset I

Using the Ames housing dataset (*www.amstat.org/publications/jse/v19n3/decock/AmesHousing.xls*), answer the following questions through a Jupyter Notebook. A description of the dataset is available at *jse.amstat.org/v19n3/decock/DataDocumentation.txt*

1. Visualize the univariate distribution of each feature, and the distribution of the target. Do you notice anything? Is there something that might require special treatment?

2. Visualize the dependency of the target on each continuous feature

3. Split the data in training and test set. Then, evaluate the ordinary least square (OLS) and visualize the relationship of the categorical variables that provide the best R^2 value with the target. Does scaling the data with *StandardScaler* help?

4. Visualize the coefficients of the resulting model. Do they agree on which features are important?

Energy Efficiency II

A multivariate linear regression model has been built to predict the **heating load** in a residential building based on a set of descriptive features, describing the characteristics of the building. Heating load is the amount of heat energy required to keep a building at a specific temperature, during winter regardless outside temperature. The descriptive features used are the overall surface area of the building, the height of the building, the area of the building's roof, and the percentage of wall area in the building that is glazed. This kind of model would be useful to architects or engineers when designing a new building.

Using the Energy efficiency dataset (*archive.ics.uci.edu/ml/datasets/energy+efficiency*), build a linear model to make predictions for each of the query instances shown in the table below.

Surface area	Height	Roof area	Glazing area
784	3.5	220.5	0.25
710.5	3.0	210.5	0.10
563.5	7.0	122.5	0.40
637.0	6.0	147.0	0.60