Regression Assignment: Predicting library occupancy

1 Task Description

The renovation work in the public library of a certain region needs to estimate the library occupancy.

The library director believes that library occupancy is influenced by the weather conditions, type of day (weekend, workday) and the library staff.

In this assignment, you will use regression techniques in order to predict the use of the library by using climatic, temporal and employees information.

The quality of your predictions will be measured according to the Mean Absolute Error (MAE) metric.

2 Dataset Description

The dataset library.csv contains an hourly information of weather conditions, workday type and library staff information. The target attribute is the library occupancy.

Number of instances: 10427

Number of attributes: 14

2.1 Target Class:

Occupancy: number of users in the library

2.2 Attribute Information:

Item	Attribute	Type	Values
1	**dteday**	Date	Date
2	**season**	Categorical	spring, summer, fall, winter
3	**hr:** hour of the day	Numerical/Categorical	0-24
4	**holiday**	Binary	0:no holiday, 1: otherwise
5	**weekday:** day if the week	Categorical	0-6
6	**workingday**	Binary	0: no working day, 1: otherwise
7	**weather**	Categorical	Clear, Cloudy, LightRain, Snow
8	**temperature**	Numerical	Degrees
9	**stemp:** temperature perception	Numerical	Degrees
10	**humidity**	Numerical	Percentage
11	**winds:** windspeed	Numerical	velocity
12	**nemployee:** average number of employees	Numerical	number
13	**expemployee:** average year of experience of employees	Numerical	years
14	**occupancy:** TARGET	Numerical	number of users in the library

3 Predictions

The prediction of your model can be evaluated in the following dataset:

- prediction_explanatory: contains the explanatory variables of 6050 independent records.
- prediction_target: contains the real value of the previous observations that you can confront with your predictions.