Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
July 12th, 2022

Problem n.3

Nancy and Jhonatan are searching an apartment to rent in Boston. For this reason they would like to understand which variables influence the `price` of the rent (in $). The file `rent.txt` contains several variables regarding the different apartments: the square `footage` (m^2), the presence of `two bathrooms`, the `age` of the building (years), the number of years since last `renovation`, the distances from public `transport`, from the city `center`, from the closer `supermarket` and from the closer `park` (in meters).

a) Formulate a linear regression model for the `price`, as a function of all the other variables. Include in the model a possible dependence on the categorical variable `two bathrooms`, both in the intercept and in the interaction with all the other numerical variables. Report the estimate for all the coefficients of the model and the errors' standard deviation.

b) Analyze the model residuals and verify the assumptions of the model.

c) Perform a variable selection through a Lasso method, by setting the parameter controlling the penalization to $\lambda = 45$. Report the significant coefficients.

d) Optimize the parameter λ within the range [1; 100] via cross-validation. Report the optimal λ and the corresponding estimated coefficients.

e) Provide pointwise prediction for the price of an apartment with `square footage` = $30m^2$, `age` = 5, `renovation` = 5, `transport` = 300m, `center` = 1000m, `supermarket` = 500m, `park` = 100m, and one bathroom.

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUMUpJUFQ1UTVCQThBUVhOM1E4U1ZWTEdSQy4u`