

Practical Work 5

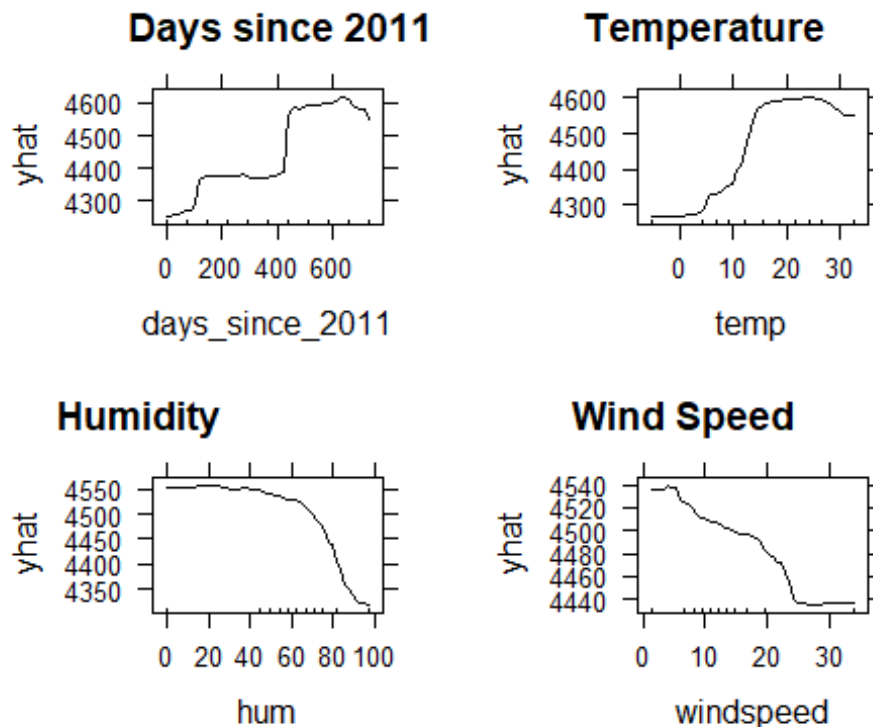
Hugo Albert and Pablo Sánchez

2023-04-26

Deployment: Model-agnostic methods

1.- One dimensional Partial Dependency Plot.

In this section, we will make use of dependency plots to understand how the a predicted feature behaves depending some of the variables that the model considers. This plot is very powerful and intuitive for explainable purposes and gives a quick insight into possible causal relationships between independent and dependent variables. In this case, our target variable will be the rented bikes of the dataset we have been using in the last practical works.



On the one hand, the number of days since 2011 and the temperature are positively related to the number of rented bikes in a particular date, both with a little decrease if the number is too high. The number of days since 2011 show a particular sharp raise around 420 days and a really sharp but smaller one around 100, remaining almost constant in between those increases. However, the temperature influence raises also really fast but gradually between 5 degrees and 15, and remains constant until it reaches 27 degrees approximately, when it starts decreasing a bit. We do not have enough temperature values to ensure that it will continue decreasing after 30 degrees, but it looks like a reasonable thought.

On the other hand, high values of humidity and wind speed seem to decrease the number of rented bikes. Both have two extreme values in the beginning and the end of the distribution which are separated from the rest of the quantiles, leaving a big gap without enough data to draw conclusions (although it seems that the wind speed distribution is symmetric to the humidity's). However, we have enough data to state that a raise of any of the variables would lead to a decrease of the number of rented bikes. Humidity remains constant for some values, but suddenly decreases relatively fast, while the decrease produced by wind speed is gradual and almost linear until it reaches the gap with few data.

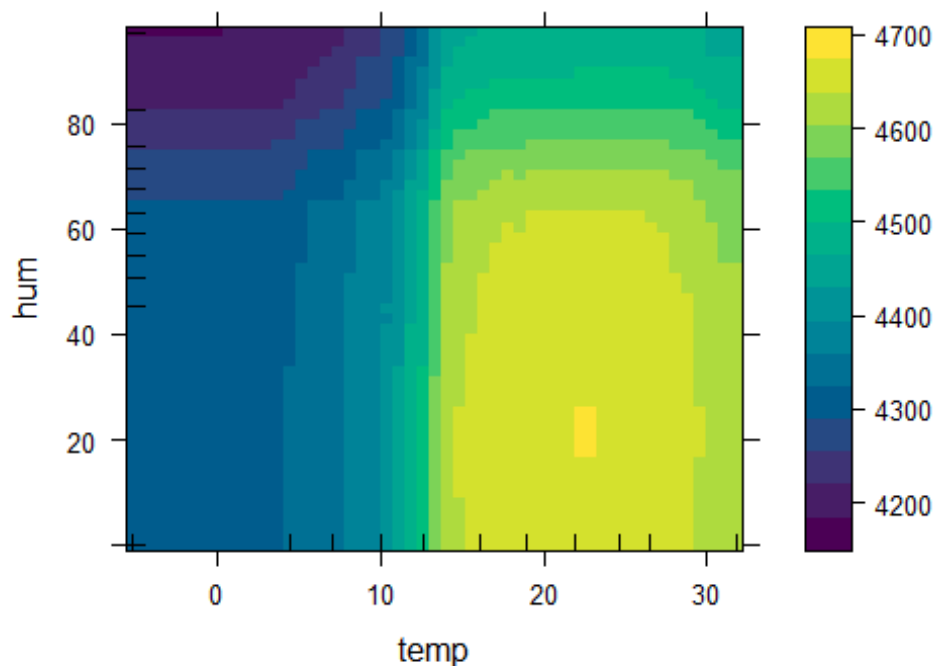
2.- Bidimensional Partial Dependency Plot.

Bidimensional partial dependency plots are usually a very straightforward way to draw conclusions about the impact that the interaction of features has on the target variable. Nonetheless, it must be taken into consideration the fact that if the features do not apply being independent can lead to not fully realistic conclusions.

This arises because for unseen combination points of variables, the model has to create observations where the actual feature distribution probability is very low. In this case, we will be analyzing how the Humidity and Temperature infer on the rented bikes as a result of their interaction. Having said this and having into account some of

the comments made in the previous section, we could expect that the interpretation might be unreliable at extreme combinations of the features such as very high temperatures together with extreme levels of humidity and in the opposite case.

This will be confirmed with the data distribution, since 1st quantile is very wide for both of the variables, what might reveal no samples at very low values and very short 10th quantile.



Our thoughts about feature distributions are apparently, so we will have to take into account the break of the independence assumptions.

Nevertheless, this bidimensional dependency plot reinforces some of the drawn conclusions from previous practical exercises. We stated out that despite the temperature effect plot revealed that the higher that temperature was, the more daily rented bikes might be expected, this was not feasible at all from a practical point of view. The same way, it could be seen that drier humidity conditions had a significant negative effect on the rented bikes.

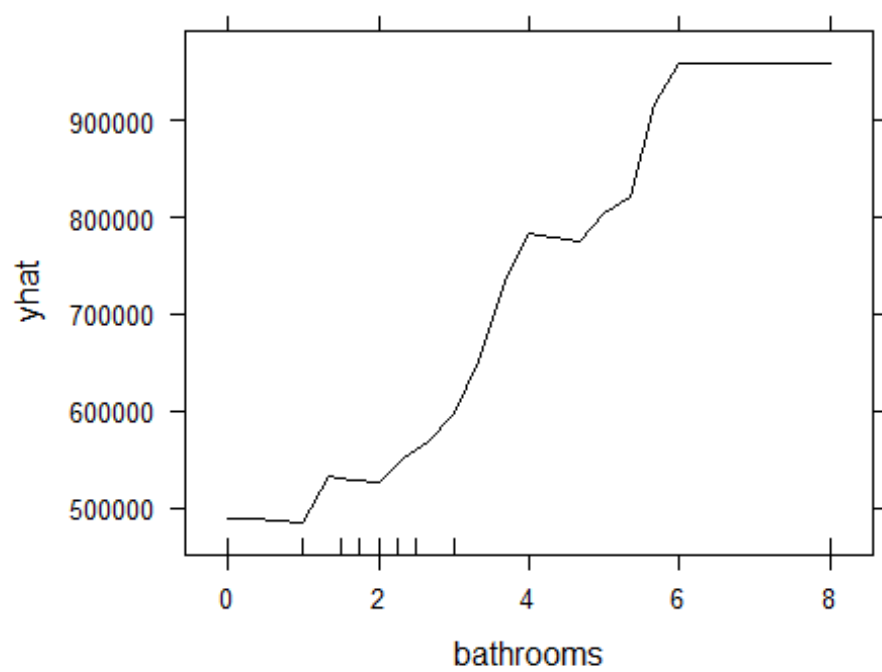
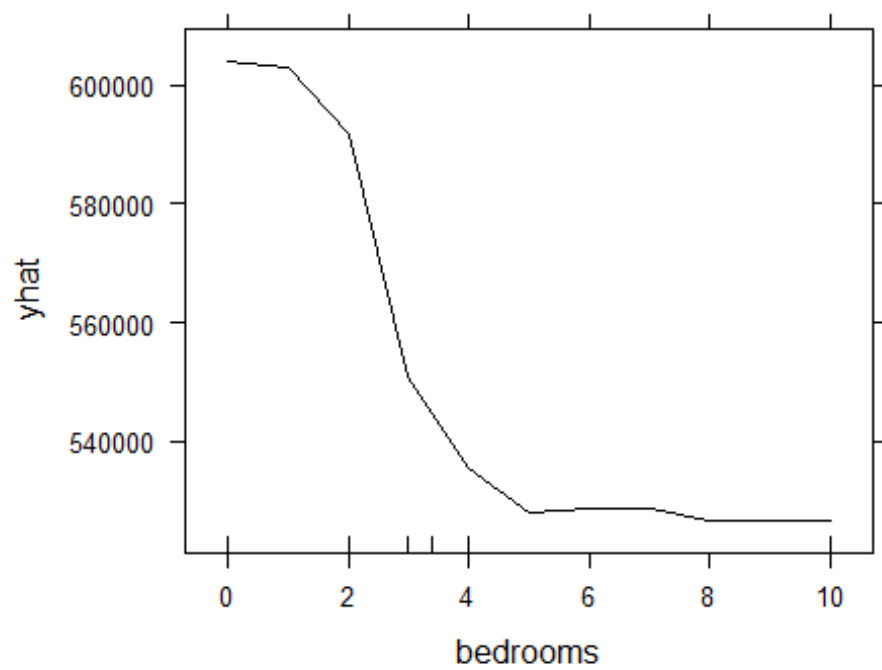
This could lead us to think that the best conditions for high bike rentals were very hot and humid days. From a realistic perspective, we might expect that smooth temperatures and slight humid levels could encourage people to rent bikes, as that weather conditions are comfortable for the user. In fact, this is what the bidimensional dependency plot reveals, since the highest predicted values are concentrated on temperature values in the 15-30 degrees interval and for humid concentrations around 25%. These conditions contradict what effect plots revealed but are more realistic.

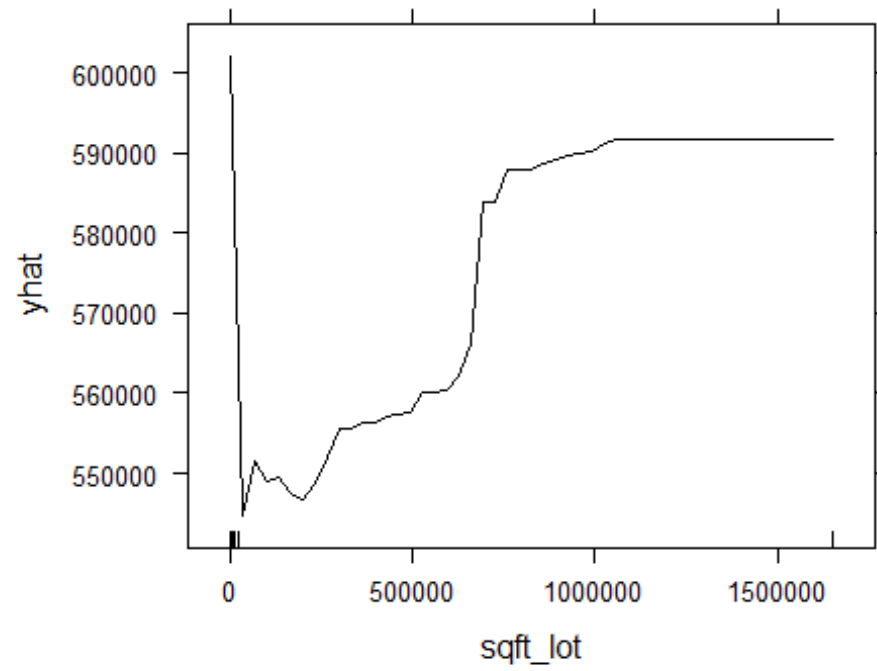
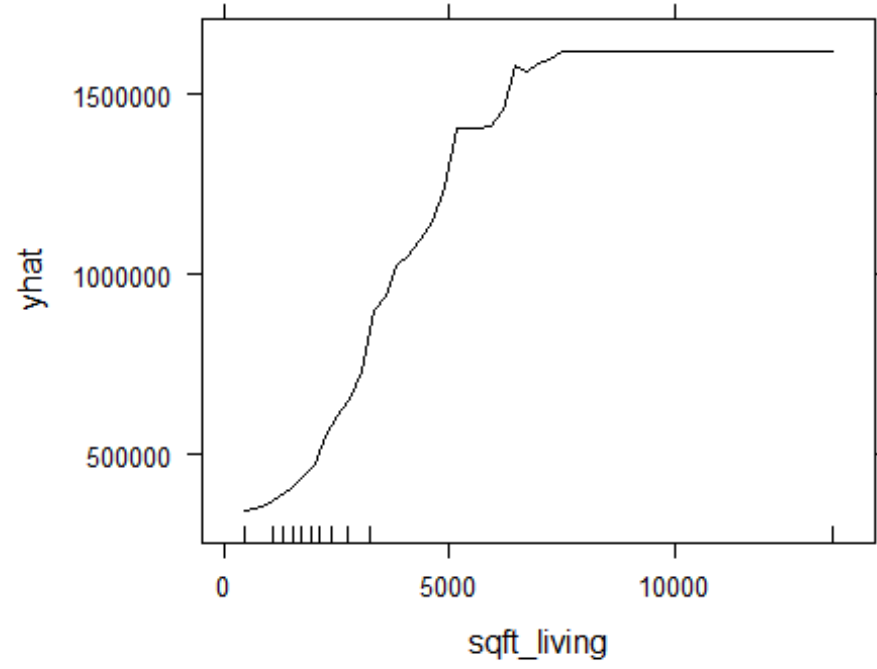
The same way, it can be seen that the conditions that imply the most negative outcome for bike rentals, revealing a very few rented bikes on that conditions were very high umidity levels and temperatures around 0 degrees.

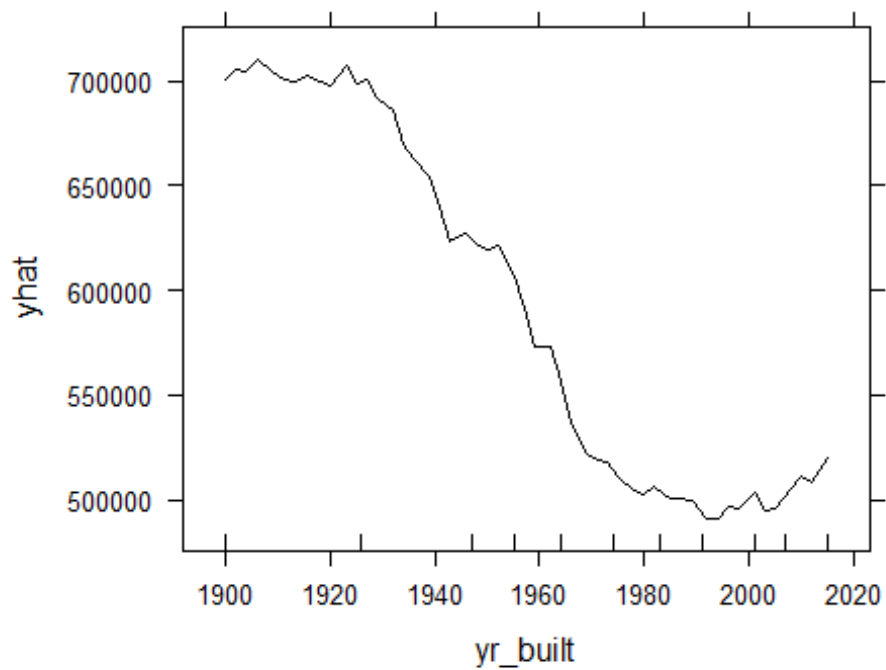
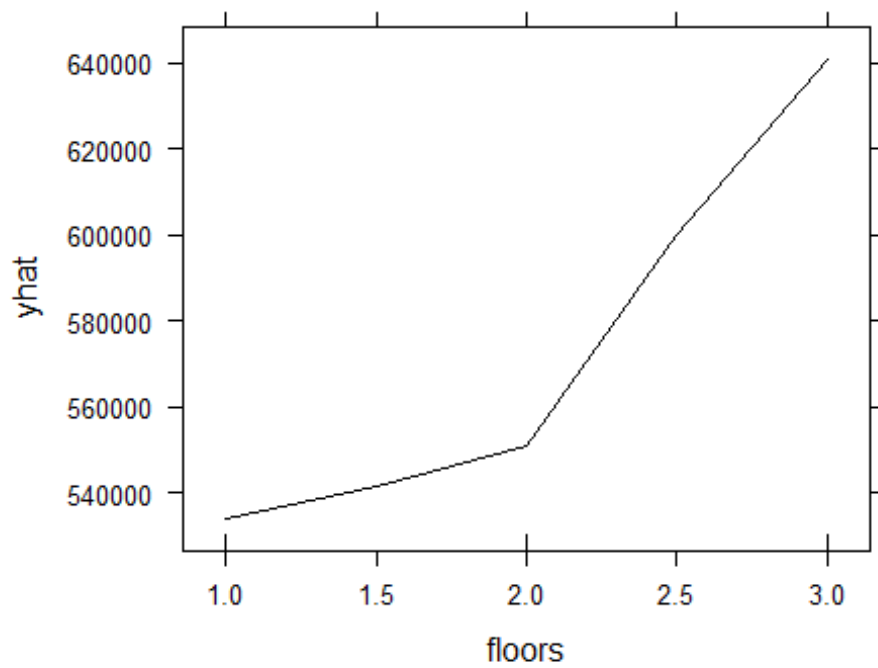
Meanwhile, in general it can be seen that temperatures under 15 degrees and humidity levels over 60 infer considerably negatively on the rented bikes.

3.- PDP to explain the price of a house.

In this case, we aim to explain how predictions on house prices depend and behave depending on some of the property characteristics from an individual point of view, and then, considering the interaction between some of them.

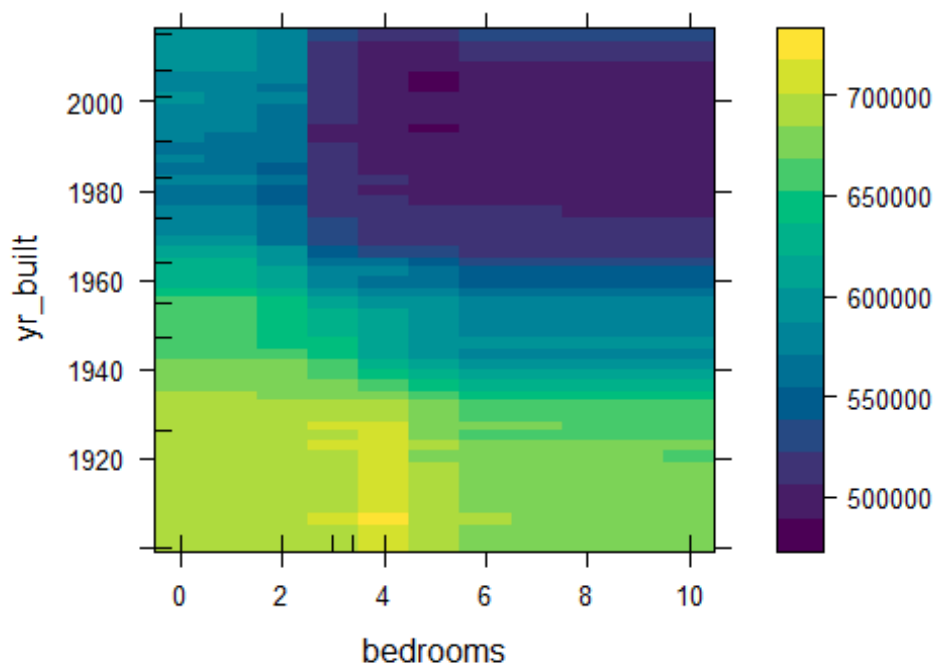






The number of bathrooms, the living square feet and the number of floors tend to elevate the price of the house, especially the number of bathrooms and the living

square feet are the ones that produce a higher difference in the price, whereas the number of bedrooms, the lot square feet and the year where it was build decrease the price of the house with their first raise and then they reach a point where they make it increase again. However the reasons may vary. The lot square feet of the house present a really asymeric distribution which may be affecting the plot, while the other two variables are following more symetric ones, so it would be interesting to show the relation between both of them.



Here we can appreciate that the old houses are more expensive but the most current ones increase their price if they have a little number of rooms or too many of them. However, it is difficult to stablish a relation between them, but there is a gap where the prices descend which is located from 1980 and 2000 around 5 and 6 bedrooms, a number of bedrooms that did not decrease the price so much in the past nor in the present (maybe because of a tendency of having more descendants during the marriage, as now families do not need so many rooms).