

Hausarbeit im Modul „Data Science und Machine Learning“ WS22/23- Teil 2

The Bike Shop Data

In that part of your homework, you should predict how many bike rentals happen for a day (*cnt*). Please use the following data and respond to the questions below.

Attribute Information:

- instant: record index
- dteday: date
- season: season (1: winter, 2: spring, 3: summer, 4: fall)
- yr: year (0: 2011, 1:2012)
- mnth: month (1 to 12)
- holiday: weather day is holiday or not
- weekday: day of the week
- workingday: if day is neither weekend nor holiday is 1, otherwise is 0.
- weathersit:
 - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
 - 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
 - 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
 - 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp: Normalized temperature in Celsius. The values are derived via $(t-t_{min})/(t_{max}-t_{min})$, $t_{min}=-8$, $t_{max}=+39$ (only in hourly scale)
- atemp: Normalized feeling temperature in Celsius. The values are derived via $(t-t_{min})/(t_{max}-t_{min})$, $t_{min}=-16$, $t_{max}=+50$ (only in hourly scale)
- hum: Normalized humidity. The values are divided to 100 (max)
- windspeed: Normalized wind speed. The values are divided to 67 (max)
- casual: count of casual users
- registered: count of registered users
- cnt: **count of total rental bikes** including both casual and registered (**AV**)

Tasks:

1. Train and evaluate two different multiple regression models using different optimization algorithms. Describe your approach, interpret the performance, and justify which machine learning model you would choose.
2. Train and evaluate a polynomial regression model. Test at least two different models. Describe your approach, interpret the performance, and compare the results to the chosen machine learning model from task 1. Which machine learning model would you choose?

3. Select the machine learning model that you have chosen in task 2. Now try optimizing your results by using hyperparameter tuning (lecture 04). Use cross-validation (lecture 04) when evaluating your results. Please explain your results and interpret them.
4. Apply an ensemble learning technique (lecture 05) to see whether this technique leads to a better performance than in tasks 1 and 2. Use hyperparameter tuning (lecture 04) if needed and use cross-validation (lecture 04) when evaluating your results. Describe your approach, interpret the performance, and justify the final machine learning selection.