Practice 6A: Non-Parametric Regression

Supervised Learning - 10 December 2021

1. Load the dataset. The dataset is accessible through `sklearn`

   ```
   from sklearn.datasets import load_diabetes
   data = load_diabetes()
   ```

2. Select one of the features f for the consecutive analysis. Split the dataset in train and test and use the training set to fit a *piecewise step function.* Evaluate the model on the test set.

3. Exploit the same feature selected above, and the same train and test sets to train and validate a *Piecewise polynomial* degree 3 with 5 knots (arbitrarily chosen).

4. Use the same data exploited above to train a *Natural Spline*, degree 3 and 5 knots. Evaluate the model on the test set, and compare the evaluation with both the piecewise step function and the piecewise polynomial trained above.

5. Is there any significant difference in the performances of the three models? Which is the reason for this difference? Visualize the data (selected feature f versus target variable) and the trained models. The performances results are also visible in this graph?