Homework

To be completed by 16 DEC, 2019

Assignment 1

- Express F1 in terms of the entries in the confusion matrix (TP, FP, TN, FN) and simplify the equation.

Assignment 2

- Learn about the derivation of the Naïve Bayes formula https://en.wikipedia.org/wiki/Naive_Bayes_classifier

$$p(C_k, x_1, \ldots, x_n) \quad = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})} \quad = \quad \ldots \quad = \quad p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

Assignment 3

- Compare the Naïve Bayes classifier with decision trees.
- How do we evaluate the Naïve Bayes classifier? Methods, metrics.
- Estimate the probabilities of C1 and C2 in the table below by relative frequency and Laplace estimate.

Število dogodkov		**Relativna frekvenca**		**Laplaceova ocena**	
tipa C1	tipa C2	P(C1)	P(C2)	P(C1)	P(C2)
0	2				
12	88				
12	988				
120	880				

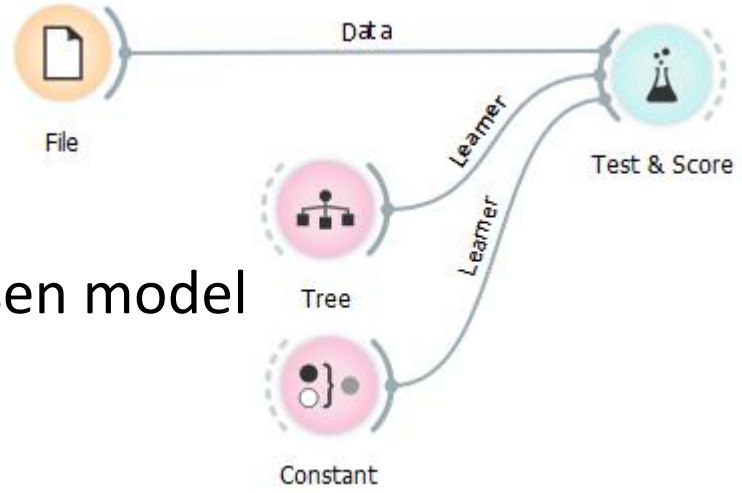
Assignment 4

- Read

Loh, Wei-Yin. "Classification and regression trees." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1.1 (2011): 14-23. https://onlinelibrary.wiley.com/doi/full/10.1002/widm.8

- Compare decision and regression trees.
- Rules of thumb when choosing the k parameter of KNN.

Assignment 5



- Use Orange and a calculator to compute RRSE for a chosen model
- Data: regressionAgeHeight.scv

- RRSE = root relative squared error
  - Nominator: sum of squared differences between the actual and the expected values
  - Denominator: sum od squared errors

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i - a_i)^2}{\sum_{i=1}^{n}(\bar{a} - a_i)^2}}$$

  - RRSE: Ratio between the error of the model and the error of the naïve model (predicting the average)
  - Hint: If we divide both the nominator and the denominator by n we get RSE of the model and const model.