

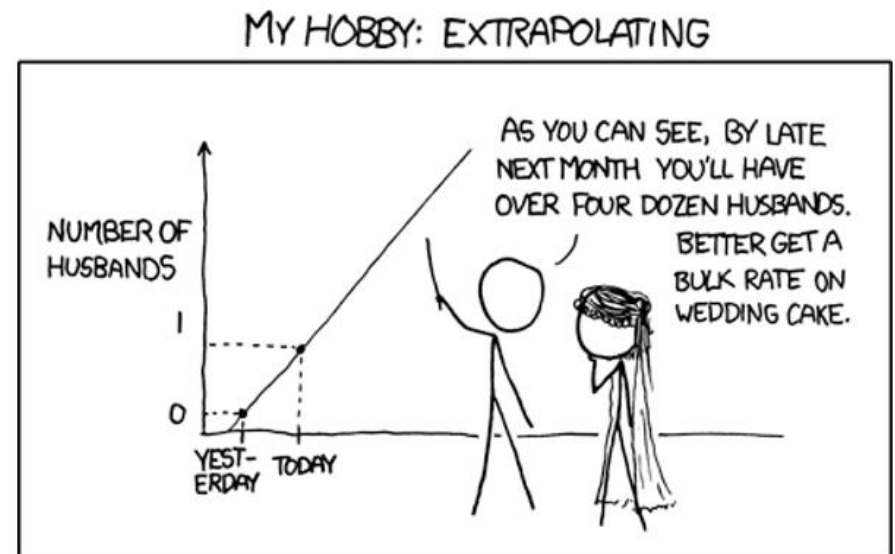
Regression

Exercise 6



Recap: Regression

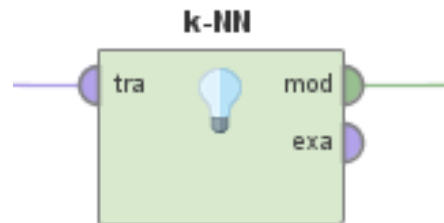
- Classification predicts a *nominal* value
 - A finite set of values
- Regression predicts a *numerical* value
 - A possibly infinite set of possible values
 - Can be *interpolating* and *extrapolating*



<http://xkcd.com/605/>

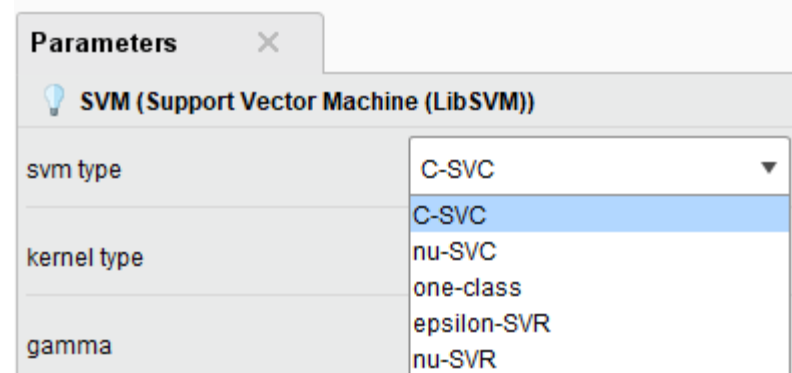
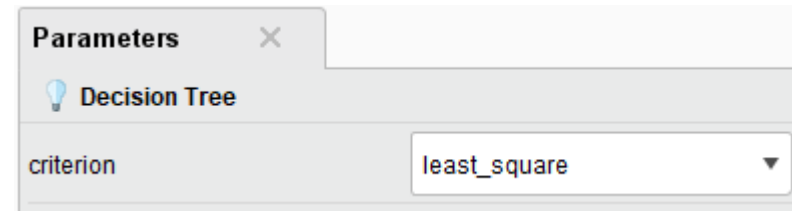
K Nearest Neighbours Regression

- Find the k nearest neighbours
- And use the average of their label as prediction
- Only interpolating regression possible
- It's the same operator that you already know from classification!



Regression Trees / SVM / ANN

- Other operators that you already know, which can be used for regression:
- Decision Tree
 - Set criterion to „least square“
- SVM
 - Set svm type to „epsilon-SVR“ or „nu-SVR“
- Neural Net
 - No changes required



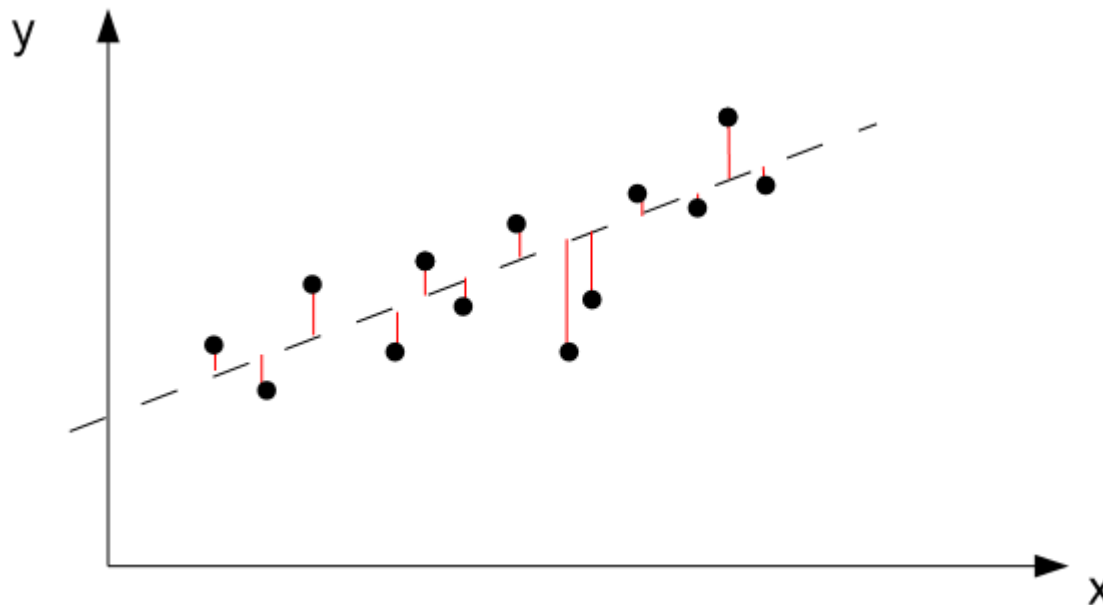
Linear Regression

- Finds a linear function

$$f(x) = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

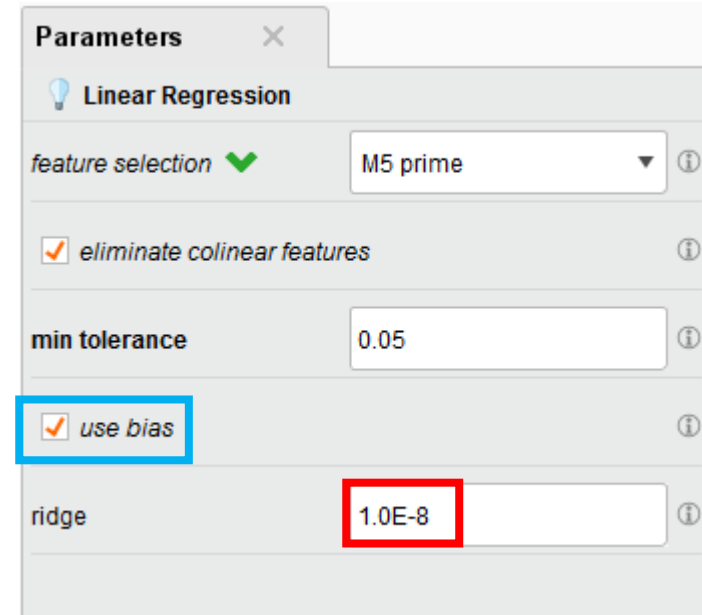
- That minimises the error

$$\sum_{\text{all examples}} (w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n - y)^2$$



Operators: Linear Regression

- A learning operator that learns a linear regression model
 - Selects the features automatically
- Parameters:
 - Feature selection:
 - none, M5 prime, greedy, T-Test, Iterative T-Test
 - Use bias:
 - determines if an intercept should be used in the regression
 - Ridge:
 - Controls the slope of the learned function, higher ridge results in smaller coefficients



Parameters

Linear Regression

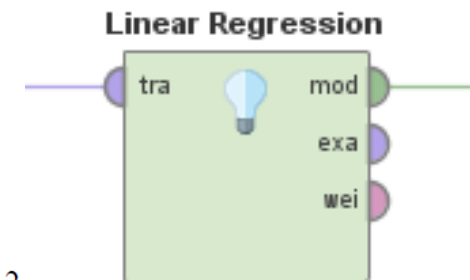
feature selection ☒ M5 prime

☒ eliminate colinear features

min tolerance 0.05

☒ use bias

ridge 1.0E-8

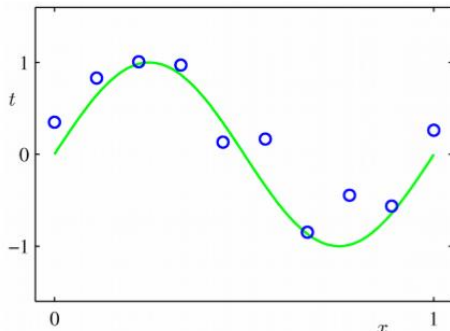


$$\sum_{\text{all examples}} \left(\underline{w_0} + w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n - y \right)^2 + \underline{\lambda} \sum_{\text{all variables}} w_i^2$$

Operators: Polynomial Regression

Original attributes are transformed before running a linear regression

- Parameters:
 - Replication factor:
 - How often can a feature be replicated in the transformation?
 - Max degree:
 - Maximal degree of the final polynomial
 - Min/Max coefficient:
 - Limit the values of the coefficients

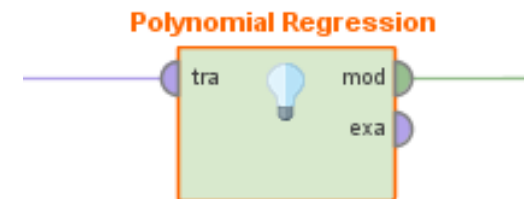


$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M$$

Parameters ×

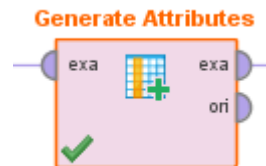
Polynomial Regression

max iterations	5000	
replication factor	1	
max degree	5	
min coefficient	-100.0	
max coefficient	100.0	
☐ use local random seed		



Operators: Polynomial Regression

- Careful: the polynomial regression operator does not always produce the expected result!
- Alternative: Manually create a polynomial regression
 - Using the generate attributes operator
 - And a linear regression afterwards



Edit Parameter List: function descriptions

Edit Parameter List: **function descriptions**
List of functions to generate.

attribute name	function expressions
age2	age*age
age3	age*age*age

Operators: Local Regression

- Lazy Learning!
- Retrieves the k nearest neighbours, calculates a regression model, then predicts the value
- Parameters:
 - Degree:
 - Degree of the locally fitted polynomial
 - Measure, neighbourhood type, k:
 - Used to select the nearest neighbours

Parameters

Local Polynomial Regression

degree 2

ridge factor 1.0E-9

☐ use robust estimation

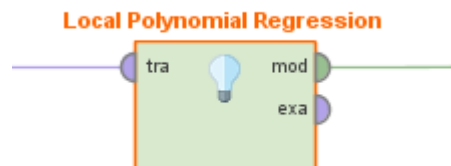
☒ use weights

numerical measure EuclideanDistance

neighborhood type Fixed Number

k 5

smoothing kernel ☒ Triweight



Performance Measures for Regression

- Mean Absolute Error

- How far are we off on average?

$$\text{MAE} = \frac{\sum_{\text{all examples}} |predicted - actual|}{N}$$

- Root Mean Squared Error

- Re-scales the errors:
 - Large errors have more influence
 - Small errors have less influence

$$\text{RMSE} = \sqrt{\frac{\sum_{\text{all examples}} |predicted - actual|^2}{N}}$$

- Coefficient of Determination (R^2)

- Tells you how much of the variation of your target variable is explained by the model

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$