Collection of exercises – Supervised Regression

Lecture exercises

Exercise 1:

Let the 2D feature vectors in the following figure be with two different numeric target values (2 and 4). Predict the point (7,6) - represented by the grey point in the picture - with the k-nearest neighbor method. Distance function should be the L_1 norm (Manhattan distance):

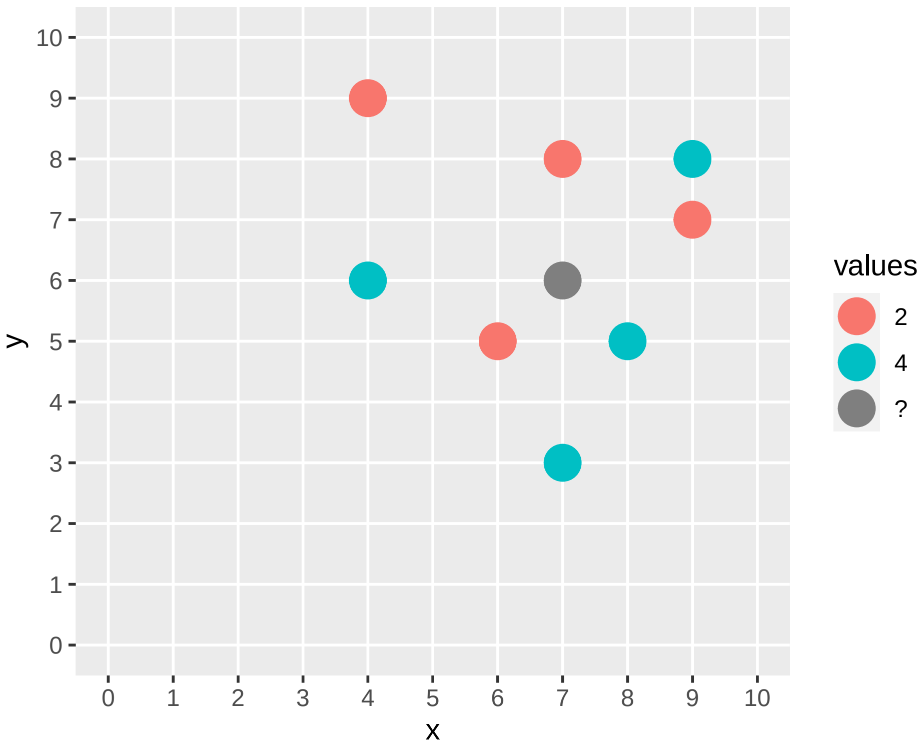
$$d_{\text{manhattan}}(x, \tilde{x}) = \sum_{j=1}^{p} |x_j - \tilde{x}_j|$$

State as the prediction the unweighted and the weighted (according to the Manhattan distance) mean of the values of the k-nearest neighbors.

a) $k = 3$

b) $k = 5$

c) $k = 7$



Exercise 2:

How in mlr3 a learner can be constructed and what it represents can be found at `https://mlr3book.mlr-org.com/learners.html`.

a) How does a learner in mlr3 compare to what you've learned in the videos?

b) Pick an mlr3 learner of your choice. What are the different settings for this learner? (Hint: Use `mlr_learners$keys()` to see all available learners)

Exercise 3:

We want to predict the age of an abalone using its longest shell measurement and its weight.

See: `http://archive.ics.uci.edu/ml/datasets/Abalone` for more details.

a) Plot `LongestShell`, `WholeWeight` on the x- and y-axis and color points with `Rings`

Using the mlr3-package:

b) Fit a linear model

c) Fit a k-nearest-neighbors model

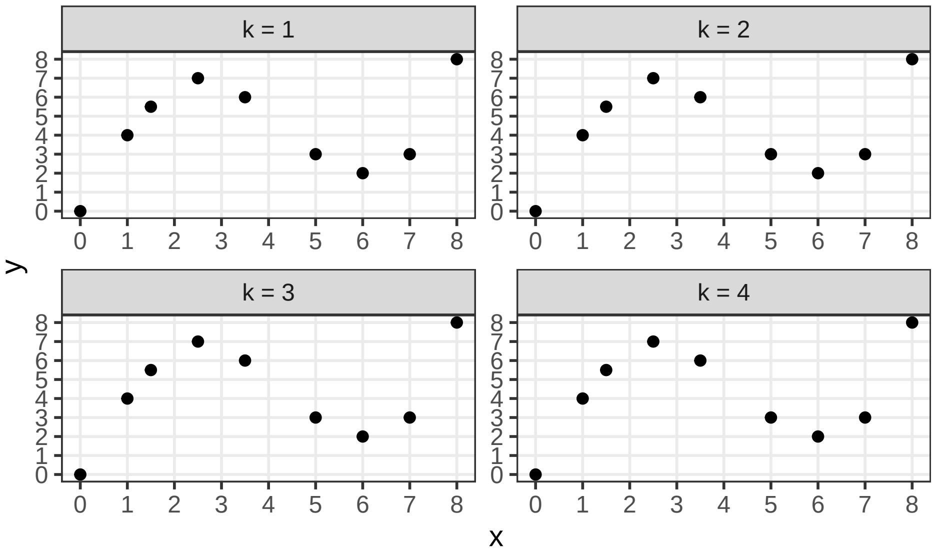
d) Compare the fitted and observed targets for lm and knn, respectively (Hint: Use `autoplot()`)

Hint: See the official book manual of the mlr3 package for usage:

`https://mlr3book.mlr-org.com/index.html`

Tasks from past exams

WS2020/21, retry exam



(a) Now we want to train a cubic polynomial, i.e., a polynomial regression model with degree $d = 3$ on the data used in a).

(i) Define the hypothesis space of this model and state explicitly how many parameters have to be estimated for training the model.

(ii) Define the minimization problem that we have to optimize in order to train the polynomial regression model. Use L2 loss and be as explicit as possible - without plugging in the data.

(iii) In order to estimate the parameters of the model, it is convenient to describe the model as a linear model. Compute the respective design matrix using the concrete values of $\mathbf{x}$ given above. Additionally, state a formula for estimating the parameters using this design matrix. (You do not have to derive this formula.)

Ideas & exercises from other sources