

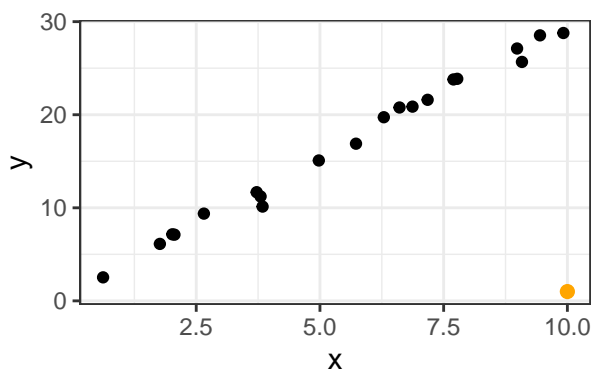
Exercise 1: HRO

Throughout the lecture, we will frequently use the R package `mlr3`, resp. the Python package `sklearn`, and its descendants, providing an integrated ecosystem for all common machine learning tasks. Let's recap the HRO principle and see how it is reflected in either `mlr3` or `sklearn`. An overview of the most important objects and their usage, illustrated with numerous examples, can be found at <https://mlr3book.ml-org.com/basics.html> and <https://scikit-learn.org/stable/index.html>.

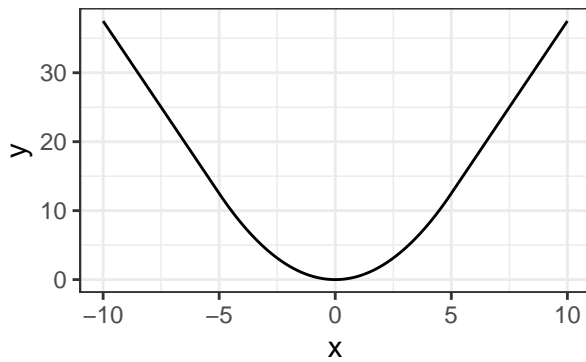
- a) How are the key concepts (i.e., hypothesis space, risk and optimization) you learned about in the lecture videos implemented?
- b) Have a look at `mlr3::tsk("iris")/from sklearn.datasets import load_iris`. What attributes does this object store?
- c) Pick a module for classification or regression of your choice. What are the different settings for this learner?
(R Hint: use `mlr3::mlr_learners$keys()` to see all available learners.)
(Python Hint: Import the specific module and use `get_params()` to see all available settings.)

Exercise 2: Loss Functions for Regression Tasks

In this exercise, we will examine loss functions for regression tasks somewhat more in depth.



- a) Consider the above linear regression task. How will the model parameters be affected by adding the new outlier point (orange) if you use
 - i) $L1$ loss
 - ii) $L2$ lossin the empirical risk? (You do not need to actually compute the parameter values.)



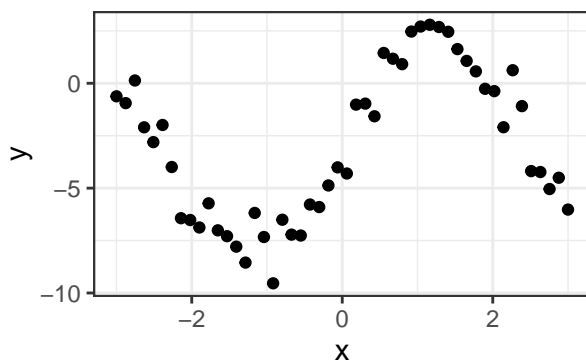
- b) The second plot visualizes another loss function popular in regression tasks, the so-called *Huber loss* (depending on $\epsilon > 0$; here: $\epsilon = 5$). Describe how the Huber loss deals with residuals as compared to $L1$ and $L2$ loss. Can you guess its definition?

Exercise 3: Polynomial Regression

Assume the following (noisy) data-generating process from which we have observed 50 realizations:

$$y = -3 + 5 \cdot \sin(0.4\pi x) + \epsilon$$

with $\epsilon \sim \mathcal{N}(0, 1)$.



- We decide to model the data with a cubic polynomial (including intercept term). State the corresponding hypothesis space.
- State the empirical risk w.r.t. θ for a member of the hypothesis space. Use $L2$ loss and be as explicit as possible.
- We can minimize this risk using gradient descent. Derive the gradient of the empirical risk w.r.t. θ . [\[Only for lecture group A\]](#)
- Using the result from d), state the calculation to update the current parameter $\theta^{[t]}$. [\[Only for lecture group A\]](#)
- You will not be able to fit the data perfectly with a cubic polynomial. Describe the advantages and disadvantages that a more flexible model class would have. Would you opt for a more flexible learner?