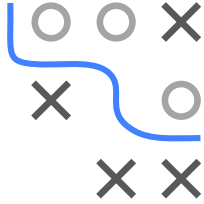
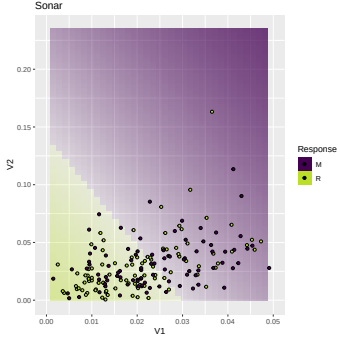
Introduction to Machine Learning



Classification: Tasks

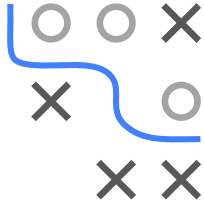


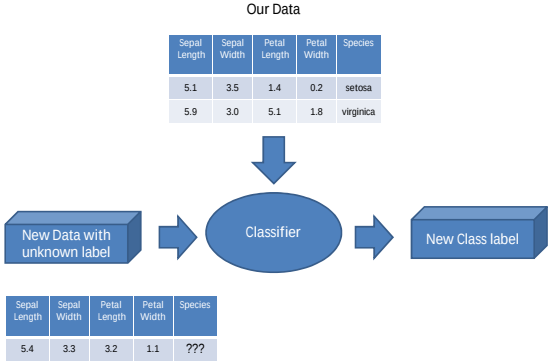
Learning goals

- Understand the main difference between regression and classification
- Know that classification can be binary or multiclass
- Know some examples of classification tasks

CLASSIFICATION

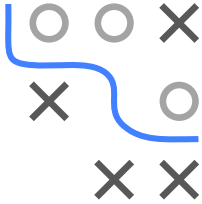
Learn functions that assign class labels to observation / feature vectors. Each observation belongs to exactly one class. The main difference to regression is the scale of the output / label.

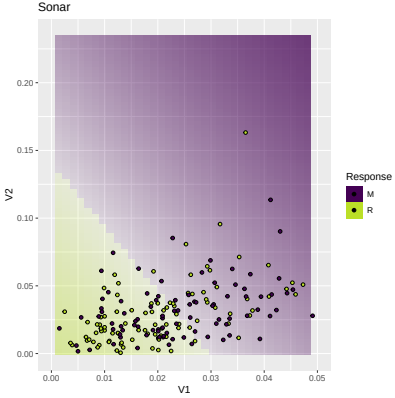




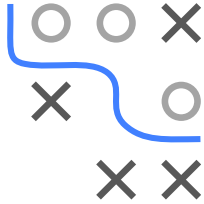
BINARY AND MULTICLASS TASKS

The task can contain 2 classes (binary) or multiple (multiclass).

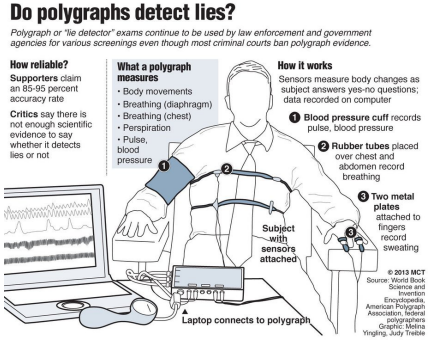




BINARY CLASSIFICATION TASK - EXAMPLES

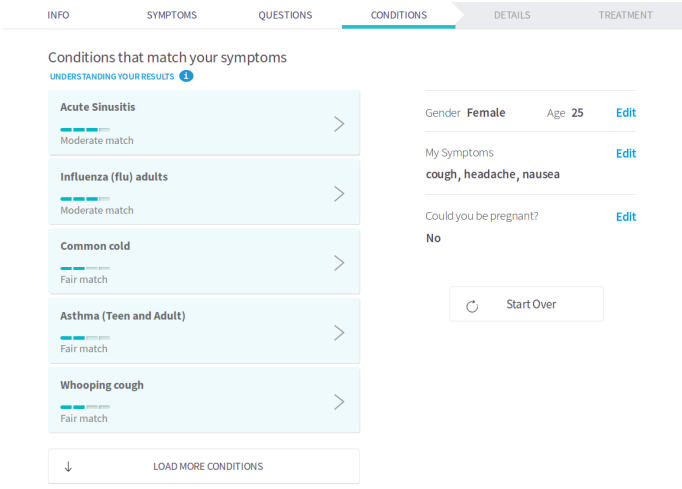


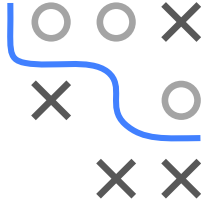
- Credit risk prediction, based on personal data and transactions
- Spam detection, based on textual features
- Churn prediction, based on customer behavior
- Predisposition for specific illness, based on genetic data



https://www.bendbulletin.com/localstate/deschutescounty/3430324-151/fact-or-fiction-polygraphs-just-an-investigative-tool

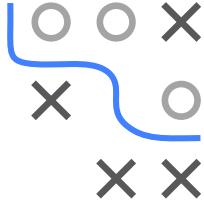
MULTICLASS TASK - MEDICAL DIAGNOSIS





https://symptoms.webmd.com

MULTICLASS TASK - IRIS



The iris dataset was introduced by the statistician Ronald Fisher and is one of the most frequent used data sets. Originally, it was designed for linear discriminant analysis.

Setosa

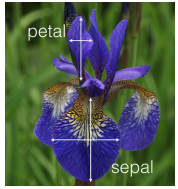
Versicolor

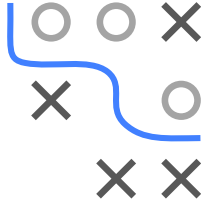
Virginica

Source:
`https://en.wikipedia.org/wiki/Iris_flower_data_set`

MULTICLASS TASK - IRIS

- 150 iris flowers
- Predict subspecies
- Based on sepal and petal length / width in [cm]





```
##       Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
##    1:          5.1         3.5          1.4         0.2    setosa
##    2:          4.9         3.0          1.4         0.2    setosa
##    3:          4.7         3.2          1.3         0.2    setosa
##    4:          4.6         3.1          1.5         0.2    setosa
##    5:          5.0         3.6          1.4         0.2    setosa
##   ---
##  146:          6.7         3.0          5.2         2.3 virginica
##  147:          6.3         2.5          5.0         1.9 virginica
##  148:          6.5         3.0          5.2         2.0 virginica
##  149:          6.2         3.4          5.4         2.3 virginica
##  150:          5.9         3.0          5.1         1.8 virginica
```

MULTICLASS TASK - IRIS