Exercise Sheet

Learning Goals

- Machine Learning Basics
- Model Formation

1. ●○○ **Machine Learning (general)** Describe the terms *supervised learning*, *unsupervised learning* and *reinforcement learning.* Sketch for each learning paradigm a typical problem together with a description of its technical realization.

   (a) Supervised Learning

   Solution: Learn a function from a set of input-output-pairs. An important branch of supervised learning is automated classification. Example: Text classification (Spam filtering): Given a dataset of emails, each annotated as *spam* or *not spam*, learn a function that classifies new emails based on word frequencies.

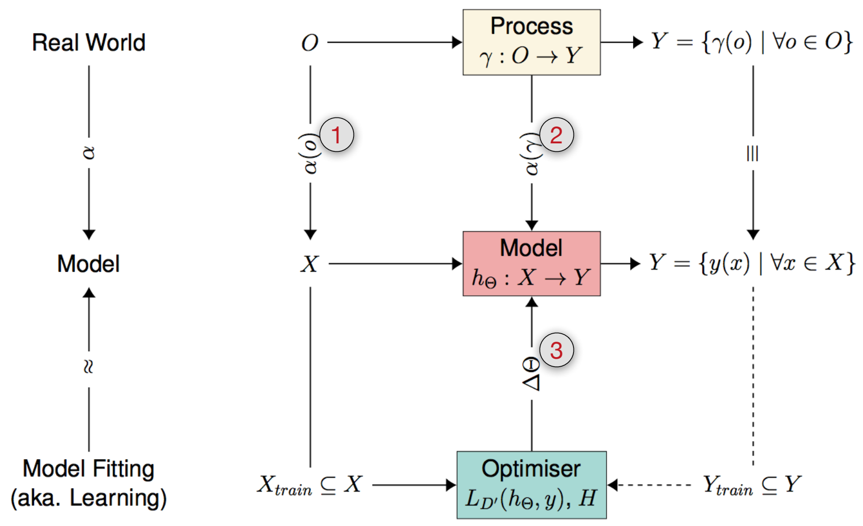
   (b) Unsupervised Learning

   Solution: Identify structures in data. Important subareas of unsupervised learning include automated categorization (e.g. via cluster analysis) and feature extraction/dimensionality reduction (e.g. factor analysis, PCA, autoencoders). Example: Cluster users by their demographic profiles and base further analysis on few clusters instead of many individual users.

   (c) Reinforcement Learning

   Solution: Learn, adapt, or optimize a behavior strategy in order to maximize the own benefit by interpreting feedback that is provided by the environment. The behavior strategy is a catalog of rules describing how to act in any possible situation. The behavior strategy is built by (a) executing sequences of actions against the environment and (b) associating individual actions with the feedback obtained at the end of the episode. Example: robot control, Alpha-Go.
   For those interested: "Reinforcement Learning". Richard S. Sutton and Andrew G. Barto. `http://incompleteideas.net/book/bookdraft2017nov5.pdf`. Very accessible exposition.

2. ●○○ **Machine Learning (general)** In the following illustration of a supervised machine learning system, describe the edges marked as (1), (2) and (3).



Solution:

1. Feature extraction
2. Choice of hypothesis space
3. Parameter update

3. **Machine Learning (general)**

(a) Which design decisions are to be made during the development of a learning system?

Solution: Example: Recognition of handwritten words in images

- Choice of performance measure; e.g. accuracy, precision, recall, etc.
- What is X and Y? e.g. Input is an array of all pixel values, output is a word.
- What is α? i.e. most likely we do not want to use the raw pixels but extract features from the image beforehand. Should we try to identify the region in the image in which the handwritten characters appear? ...or leave it to the learning algorithm to find the relevant region?
- Questions about the labels from Y
  - What if there are several words in the image? Do we receive a sequence of words as labels?
  - Do we have images showing all characters in handwriting. All pairs of characters in handwriting? Which character sequences might be hard to discern from similar looking ones?
- Determine a Machine Learning Algorithm (aka choosing the hypothesis space)

(b) Name an example of a problem which cannot be solved by learning. Explain your answer.

> **Solution:** Win a lottery.

(c) Describe the bias-variance trade-off.

> **Solution:** Any model has some a-priori assumptions built into it. The assumptions might be encoded explicitly by the analytical form of the model function or encoded implicitly by the use of a certain distance metric. The space of learnable hypotheses is restricted by these assumptions (e.g. fitting a line to data that was generated by a process with a quadratic input-output relationship won't be a good fit). If we enlarge the hypothesis space, e.g. by allowing polynomials of degree 2, we introduce more degrees of freedom (i.e. more parameters) and can better fit the data. However, if the real phenomenon has a linear relationship, and we just happen to observe a set of samples that suggest a quadratic relationship (e.g. due to outliers), our model will fit the data points in the sample. Next time, when we draw another random sample, the alleged quadratic relationship might not be present and the estimated parameters of the model end up having very different values. If we had sticked to the (smaller) hypothesis space of polynomials of degree 1, our fit on the training data might not be perfect (high bias), however, the test errors on random samples might not vary that much (low variance). A high-bias model is rather ignorant to the pecularities of the data we fit it on (underfitting).

(d) What is the key message of the ”no free lunch theorem”?

> **Solution:** There is no single algorithm, that is on average better than another algorithm over all data distributions and target functions

(e) Describe two learning problems for which two different hypothesis spaces need to be developed. The hypothesis spaces have to differ in the strength of the inductive bias (the set of the a-priori assumptions, that are made for the hypotheses).

> **Solution:** Predict the DAX based on historic stock exchange data. A-priori assumptions for the construction of the hypotheses: sine curve, spline, strict monotony.
>
> Prediction of a disease based on the data of a patient's blood picture. A-priori assumptions for the construction of the hypothesis: a single value is out of the normal region, half of the values are out of the normal region.