

# Summary of FK7068 Machine Learning for Physicists and Astronomers

Yashar Honarmandi  
yasharh@kth.se

3 november 2021

**Sammanfattning**

## Innehåll

# 1 Basic Concepts

**Machine Learning** Machine learning is the field of study that gives computers the ability to learn without explicit programming. A computer is said to learn from some experience  $E$  with respect to a task  $T$  and a performance measure  $P$  if its performance at  $T$  as measured by  $P$  increases with  $E$ .

**Supervised Learning** Supervised learning deals with labelled data, and the task is to correctly label unfamiliar data. The two main types of supervised learning problems are classification and regression.

**Unsupervised Learning** Unsupervised learning, by contrast, deals with data that does not have labels. The task is generally to identify some patterns within the data. Examples of such problems include clustering, anomaly detection, pattern recognition and association mining.

**Reinforcement Learning** Reinforcement learning is, loosely speaking, based on performing random explorations of prediction processes and learning from the output.

**A Note on Hardware** Machine learning benefits strongly from parallelization. This is one of the reasons why, despite CPUs having high clock speeds, GPUs are the preferred hardware for training algorithms.

**Using Machine Learning** Machine learning is notoriously full of pitfalls. Great care must be taken in the preparatory work and the process of creating and training the model. The general steps in the procedure are

1. **Define the problem and plan.** Formulate good questions and assess whether a machine learning approach is adequate and necessary.
2. **Estimate the required computational resources.** Combined with economic aspects, this will determine the limits of the complexity of your model.
3. **Prepare data.** This step is crucial and will include identifying sources, data collection and preprocessing. Relevant questions to ask include whether you have enough data (a quintessential point) and whether the data is sufficiently diverse.
4. **Construct an appropriate model.**
5. **Test the model.**
6. **Deploy.**

See also this article for a well-known and detailed example of how one should think.