## FEATURES AND PREPROCESSING

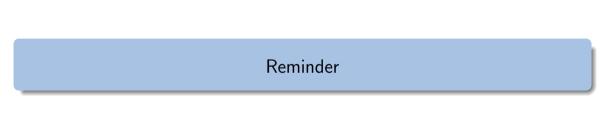
Machine Learning for Autonomous Robots

Marc Otto Robotics Innovation Center, DFKI GmbH

October 30, 2023 - Bremen, Deutschland







Exercise Sheet

The first exercise sheet was released yesterday, and is due on Wednesday in two weeks (8th of November, eod).



## Recap

- What are features? What types of features were discussed in the lecture?
- ▶ What is one-hot encoding and why is it useful?
- Explain data preprocessing in the context of Machine Learning.
- ► What is feature generation?
- What is feature selection?
- What is dimensionality reduction?
- Why are we interested in reducing the number of features?
- Explain Principal Component Analysis.

Task

 $\label{lem:control_power_power} Download\ the\ Jupyter\ notebook\ from\ StudIP\ and\ solve\ the\ tasks\ described\ there.$ 

## A Naive Feature Selector

```
NaiveFS(\mathcal{F}, \mathcal{D}, k):
      /* Selects k features from \mathcal{F} that have the
            strongest correlation with the class on training
            data \mathcal{D}
                                                                                                              */
      \mathcal{F}' \leftarrow \{\}:
      while |\mathcal{F}'| < k do
           feature \leftarrow \arg\max_{f \in \mathcal{F}} |r_{\mathcal{D}}(f, c)|;
       \mathcal{F}' \leftarrow \mathcal{F}' \cup feature;
\mathcal{F} \leftarrow \mathcal{F} \setminus feature;
      end
      return \mathcal{F}':
```

## Greedy Hill-climbing Heuristic

```
GreedyCFS(\mathcal{F}, \mathcal{D}):
      /* Selects feature subset \mathcal{F}' \subseteq \mathcal{F} that maximizes merit(\mathcal{F}') on \mathcal{D}. Uses the
             greedy hill climbing heuristic.
                                                                                                                                                                         */
      \mathcal{F}' \leftarrow \{\}:
      while True do
             last\_merit \leftarrow merit(\mathcal{F}');
             f_{sel} \leftarrow \arg\max_{f \in \mathcal{F}} merit(\mathcal{F}' \cup \{f\});
             if merit(\mathcal{F}' \cup \{f_{sel}\}) < last\_merit then
                   return \mathcal{F}':
             end
            \mathcal{F}' \leftarrow \mathcal{F}' \cup \{f_{sel}\};
\mathcal{F} \leftarrow \mathcal{F} \setminus f_{sel};
      end
```