

DEAL Python course IV: machine learning

April 11, 2023

ML principles and caveats

- DISCLAIMER: This is an implementation course, not a theory course
- ML works best with lots of data, especially lots of high-quality data
 - "Garbage in \rightarrow garbage out"
- Overfitting is one of the biggest challenges, especially with very large models
- ML is still mostly a black box (and therefore generally not suitable for causal inference)
- Training large models with lots of data quickly becomes computationally expensive
 - Potential solutions: Acquiring a GPU, renting additional computing resources from AWS