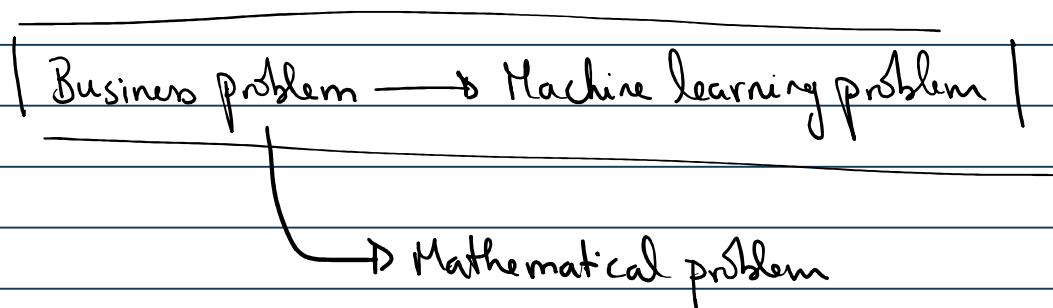


Framing the problem

Monday, 4 March 2024 6:27 pm



Type of problem → Big picture



Current solution



Getting the data [Design Features]



Metrics to measure



Online vs Batch



Check assumptions

* Embracing Ambiguity is part of a Data Scientist

- ① **Define the problem.** What is the task we want to teach a computer to do?
- ① **Collect data.** Gather data for training and testing sets. The larger and more diverse the data the better.
- ② **Design features.** What kind of features best describes the data?
- ③ **Train the model.** Tune the parameters of an appropriate model on the training data using numerical optimization.
- ④ **Test the model.** Evaluate the performance of the trained model on the testing data. If the results of this evaluation are poor, re-think the particular features used and gather more data if possible.