

10 Questions to Verify when Designing Machine Learning Experiments

- 1) What is the variable the machine learner should predict? What is the required accuracy for success? What impact will adversarial examples have?
- 2) How much data do we have to train the prediction of the variable? Are the classes balanced? How many modalities could be exploited in the data? Is there temporal information? How much noise are we expecting? Do you expect bias?
- 3) How well is the data annotated (anecdotally)? What is the annotator agreement (measured)?
- 4) Given questions 1-3: Are we reducing information (pattern matching) or do we need to infer information (statistical machine learner)? As a consequence, what seems the best choice for the type of machine learner per modality?

Per modality:

- 5) Estimate the memory equivalent capacity needed for the machine learner of your choice. What is the expected generalization? How does the progression look like.
- 6) Train your machine learner for accuracy at memory equivalent capacity. Can you reach near 100% memorization? If not, why (diagnose)?
- 7) Train your machine learner for generalization: Plot the accuracy/capacity curve. What is the expected accuracy and generalization ratio at the point you decided to stop? Do you need to try a different machine learner (if so, redo from 5)? Should you extract features (if so, redo from 5)?
- 8) How well did your generalization prediction hold on the independent test data? Explain results. How confident are you in the results?
- 9) How do you combine the models of the modalities? Explain your choice. How confident are you in the combination results (ie., does it make sense to combine)?
- 10) What are the final combined results of the system? Are the experiments documented and repeatable (if not, please make sure they are, even for bad results)? Are the experiments reproducible (speculate)?