Small Exercises 3

Parameter Inference

Problem 1: Explain the importance of the i.i.d. assumption on data for parameter inference in 140 characters or less.

IID: Making inference feasible (independence) by abstracting from a sequence of outcomes to its generating process (identical distribution).

(This explanation has 140 characters.)

Problem 2: Why would you prefer MAP over MLE?

An incomplete list of cases:

- As we saw on a very simple example: MLE can easily overfit if too little data is available.
- MAP can incorporate prior knowledge and prefer certain subjectively more credible solutions over the MLE solution—*inductive bias*.

Problem 3: Why would you prefer MLE over MAP?

An incomplete list of cases:

- MAP is not always as nicely tractable as in our lecture examples.
- In a very big data regime, prior information is less important. MLE may do the trick.

  In our coin example, if we have several million coin tosses, we can be fairly certain of our MLE estimate.

  However, this has to be applied with care: If we are dealing with estimating the parameters of a very deep neural network, which typically have several million and more parameters, the amount of data needed is exponentially higher!
- MAP can introduce *inductive bias*, i.e., by applying MAP, we prefer certain solutions over others. Often, that is why MAP is applied, but it can also lead to unwanted effects.

Problem 4: What is the fundamental advantage of the fully Bayesian approach over MLE and MAP?

Other than MLE and MAP, the fully Bayesian approach does not compute a point estimate of the parameters θ, but integrates out all effects of this parameter. Ideally, if the prior is chosen properly, this can capture the full model characteristics.