PHY 835: Exercise 2
Released Feb 9, 2021;Target Due Date: Feb 23, 2021

1. Bias vs variance trade-off

In the lectures we have discussed two schematic plots:

- Number of data points vs. Error.
- Model complexity vs. Error.

Perform an experiment with polynomials which shows this behavior.

2. Gradient descent – ADAM

In this exercise you will implement the Adam optimizer and examine its behavior on the function:

$$(1.5-x+xy)^2+(2.25-x+xy^2)^2+(2.625-x+xy^3)^2$$

Similar to the discussion in the lecture, implement the ADAM optimizer and visualize its behavior for two different starting points and two different learning rates.

For a learning rate of 10^{-3} ($\gamma = 0.9$, $\beta = 0.99$, $\varepsilon = 10^{-8}$), scan over a sensible grid of starting points. Highlight the different endpoints associated to these starting points in a contour plot. Qualitatively, describe the results you find.

3. Some Statistics

- Given a partition function (e.g. think about the 2D Ising model), how do you obtain, in principle the probability of a given field configuration?
- List a couple of algorithms on how to generate samples drawn from such an underlying probability distribution.
- How do you ensure that samples you have generated are independently drawn from the probability distribution?
- (Optional) How can you check whether two datasets are drawn from the same underlying probability distribution?