IAML - Study Guide - Week 3

Sambit Paul, Pavlos Andreadis, Nigel Goddard

January 2022

1 Introduction

Week 3 of the course focusses on the *generalisation* of models over varied data and the *evaluation* of models. This involves using techniques to improve the overall reliability of our models and assessing the performance.

2 Generalisation & Evaluation

- Generalisation refers to the reliability of the model's outcomes. Simply put, the more generalised a model is, the more it can correctly predict unseen data.
  To understand the concept of generalisability, the bias-variance trade-off is a key concept that needs to be clarified. You can refer to this article for a quick refresher. Hastie et al. [2009] Chapter 7, Sections 7.2 and 7.3 provides a good explanation of the bias-variance relationship.
- **Over-fitting** & **Under-fitting**: When a model is not well generalised, it is implicitly understood that the model is either over-fitted to the training data, or under-fitted to the problem. A visual representation is provided in Figure 1
  - Over-fitting: Fits very well to the training data, but performs poorly on unseen data. Usually happens when the model has high variance and low bias.
  - Under-fitting: Cannot fit well to the training data itself. Usually happens when the model has high bias and low variance.
- The dataset used for machine learning problems is split for 3 purposed:
  - Training set: Used to construct the model.
  - Validation set: Used to determine the hyperparameters of the model.
  - Test set: Used to estimate the generalisability of the model.

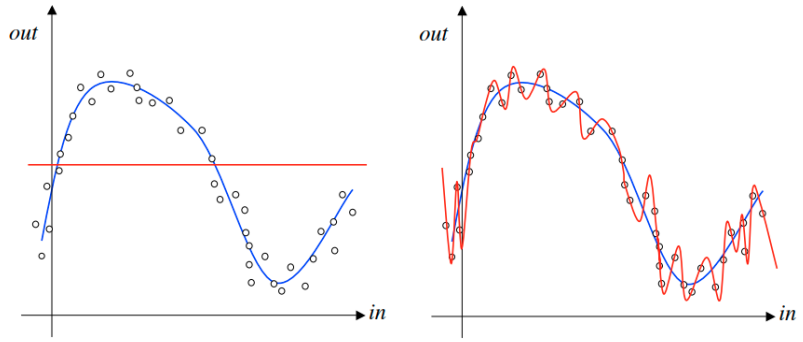

Figure 1: The left plot shows (in red) how an under-fitted model would perform, and the right plot shows (in red) how an over-fitted model would perform. The line in blue is the representation of the function.

Please refer to Hastie et al. [2009] Chapter 7, Page 222 to read about dataset splitting. In case of problems, which are not data-rich, we use cross validation which you can read about Hastie et al. [2009] Section 7.10.

- Model evaluation is done in the form of metrics. To know more about such metrics, you can refer to this article.
- Reading List:
  Goodfellow et al. [2016] pp. 105 - 117
  Witten et al. [2011] pp. 31 - 35

References

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

Ian H Witten, Eibe Frank, and Mark A Hall. Data mining: Practical machine learning tools and techniques, 2011.