Epilogue: Where to go from here

AND SO WE HAVE come to the end of our journey through the 'making sense of data' landscape. We have seen how machine learning can build models from features for solving tasks involving data. We have seen how models can be predictive or descriptive; learning can be supervised or unsupervised; and models can be logical, geometric, probabilistic or ensembles of such models. Now that I have equipped you with the basic concepts to understand the literature, there is a whole world out there for you to explore. So it is only natural for me to leave you with a few pointers to areas you may want to learn about next.

One thing that we have often assumed in the book is that the data comes in a form suitable for the task at hand. For example, if the task is to label e-mails we conveniently learn a classifier from data in the form of labelled e-mails. For tasks such as class probability estimation I introduced the output space (for the model) as separate from the label space (for the data) because the model outputs (class probability estimates) are not directly observable in the data and have to be reconstructed. An area where the distinction between data and model output is much more pronounced is *reinforcement learning*. Imagine you want to learn how to be a good chess player. This could be viewed as a classification task, but then you require a teacher to score every move. What happens in practical situations is that every now and then you receive a reward or a punishment – e.g., winning the game, or losing one of your pieces. The challenge is then to assign credit or blame to individual moves that led to such rewards or punishments being incurred. Reinforcement learning is a principled way to learn policies for deciding which action to take in which situation or state. This is currently one of the

https://doi.org/10.1017/CBO9780511973000.015 Published online by Cambridge University Press

most active subfields of machine learning. The standard reference is Sutton and Barto (1998), and you should have no trouble finding more recent workshop proceedings or journal special issues.

There are many other tasks that require us to relax some of our assumptions. For example, in multi-class classification we assume that classes are mutually exclusive. In *multi-label classification* we drop that assumption, so that an instance can be labelled with an arbitrary subset of labels. This is natural, e.g., when tagging online material such as blog posts. The dependence between labels is an additional source of information: for example, knowing that the tag 'machine learning' applies makes the tag 'reinforcement learning' more likely. Multi-label learning aims to exploit this information by learning the dependence between the labels as well as the mapping between the features and each individual label. For relevant work in the area see, e.g., Tsoumakas *et al.* (2012). A related area is *preference learning*, where the goal is to learn instance-dependent preferences between class labels (Fürnkranz and Hüllermeier, 2010). Increasing the complexity of the model outputs even further, we arrive at the general area of *structured output prediction* (Bakir *et al.*, 2007).

Going back to multi-label learning, although each label establishes a separate binary classification task, the goal is to avoid learning completely separate models for each task. This is, in fact, a special case of what is called *multi-task learning*. For example, each task could be to predict a separate real-valued target variable on the same instance space, and the learner is aiming to exploit, say, correlations between the target variables. Closely related to this is the area of *transfer learning*, which studies the transfer of models between tasks. A relevant reference for both areas is Silver and Bennett (2008).

Another assumption that deserves closer scrutiny is the availability of data in a single batch. In *online learning*, also called incremental learning, the model needs to be updated each time a new data point arrives. One application of this is in the area of *sequence prediction* (Cesa-Bianchi and Lugosi, 2006). With the increase in sensor data this setting is rapidly gaining importance, as can be witnessed from the growing area of *learning from data streams* (Gama and Gaber, 2007). Sometimes it is convenient to give the learner a more active role in data acquisition, for example by issuing queries for examples to be labelled by the teacher. *Active learning* studies exactly this setting (Settles, 2011).

Ultimately, machine learning is – and, in all likelihood, will remain – a research area at the nexus of two distinct developments. On the one hand, it is widely recognised that the ability for learning and self-training is necessary for achieving machine intelligence in any form. An area in machine learning that has this quest at heart is *deep learning*, which aims at employing hierarchies of autonomously constructed features (Bengio, 2009). On the other hand, machine learning is an indispensable tool for dealing with the data deluge. Building machine learning models is an essential step in the *data mining* process, which poses specific challenges such as being able to deal with ‘big data’ and cloud computing platforms. I hope that this book has kindled your interest in one of these exciting developments.

https://doi.org/10.1017/CBO9780511973000.015 Published online by Cambridge University Press