1 SYLLABUS
1. The Elements of Machine Learning

An Introduction to Machine Learning

Gholamreza Haffari
Updated 30 August 2018

The general problem of *Statistical Learning* is *learning* a *functional* relationship between a set of *attribute (or input) variables* and the associated *response or target variables*. The purpose of this process can be either *prediction* or *inference.* For prediction, which is the main focus of *Machine Learning*, the learned model is used to *predict* the target/response value for any (possibly new) values of the attribute variables. Whereas in inference, the models are used to *understand* the way that the target variable is affected as the input variables change.

For example, consider Figure **1.1.1** which shows the gold medal winning time for the men's 100m at each of the Olympics Games held since 1896. Our aim is to use this data to *learn* a model of the functional dependence (if one exists) between Olympics year and 100m winning time, and use this model to make *predictions* about the winning times in *future* games.

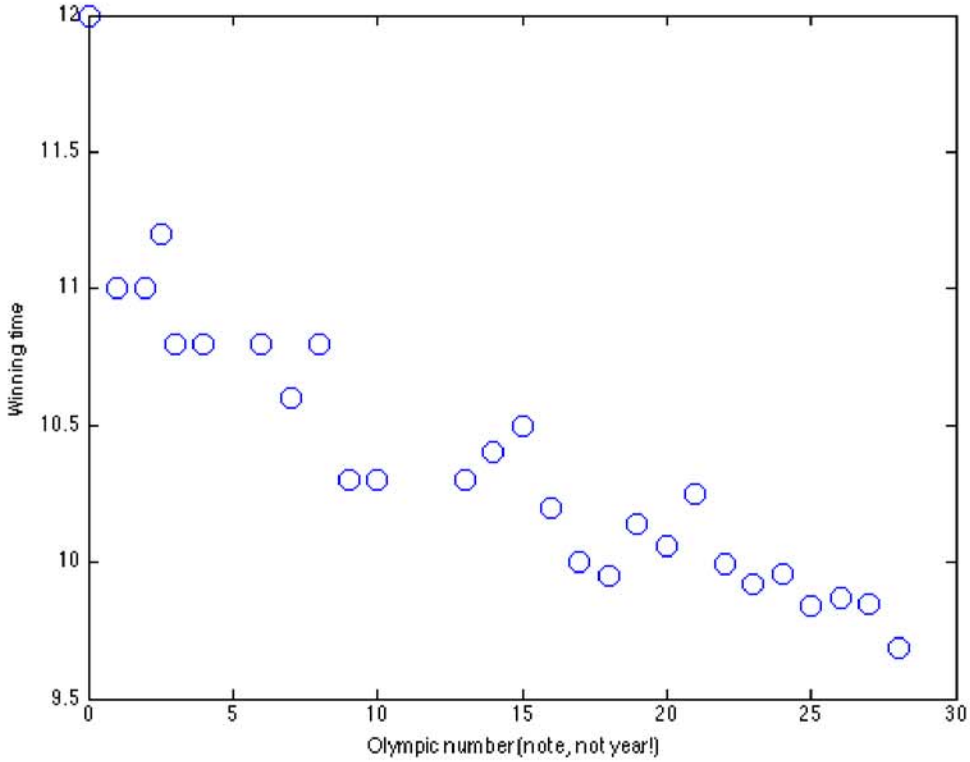

Figure 1.1.1: Winning men's 100m times at the Summer Olympics since 1896.

Clearly, the year is not the only factor that affects the winning time, but by examining Figure **1.1.1**, we can see that there is, at least, a *statistical dependence* between year and winning time (it may not be a *casual dependence* though — elapsing years are not directly causing the drop in winning times).

There are many *functions* that can be used to define the mapping between the year x (input variable) and the winning time t (target variable); mathematically, we write this as $t = f(x)$. For instance, we may consider a degree M polynomial as our function (or model):

$$f(x) := w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

where x^j denotes x raised to the power of j, and $\mathbf{w} := (w_0, \ldots, w_M)$ are the model parameters which need to be *determined somehow*. In machine learning, we *learn* the model parameters from a suitably given dataset, which is called the *training set*. For our Olympics problem, we learn the parameters based on the training data that we have collected for the past Olympic years$\{x_1, \ldots, x_N\}$ and the corresponding winning times for men's 100m.

Once the model is *trained* (e.g. its parameters are learned), it can then predict the winning time for new Olympic years, which are said to comprise a *test set*. The ability to predict accurately the target for new examples that differ from those used in the training set is known as *generalization. The holy grail of machine learning is to build models which can generalize well to unseen examples.*

Applications in which the training data comprises examples of the input variable along with their corresponding target variable are known as *supervised learning* problems. Cases such as the

Olympic winning time prediction, where the desired output is *real-valued* and *continuous*, is called *regression*. If the desired output consists of *a finite number of discrete categories*, then the task is called *classification*. An example of a classification task is to predict whether a given image includes a human face or not, i.e. the target values are 'YES' or 'NO'.

In other machine learning problems, the training data consists of a set of inputs without any corresponding target value. This is then called *unsupervised learning* problems. The goal in such problems may be to discover groups of similar examples within the data, where is called *clustering*, or to project the data from high-dimensional space down to two or three dimensions for the purpose of *visualization*.

MONASH University | Alexandria BETA