Classification

What is classification all about? Well, what we mean when we talk about classification is the process of applying categories to items in our data. By its very nature this task is very common amongst us humans. All day our brains fire constantly to decide which clothes to put on, what food to have during the day, among many other different discrete decisions.

This task entails taking data and applying a Machine Learning model to it and then with this model, draw predictions about the data. The final result leads to us taking a instance of data and deciding what category it belongs to. Typically, classification comes in two flavors: Binary (exactly 2) or Multi-class (2 or more). This distinction is a large factor in what model to use in a classification problem.

To get a better example of why classification exists in this context here is an example, suppose you are a bank and a customer walks through the door and asks for a loan. To assess whether this individual will be able to meet the payments for the loan, you as the loan representative must analyze their past fiscal behavior and through some decision-making process either provide the customer with the loan or not. Here our instance is a customer and the classification task is either to give the loan or not. The task of Classification in Machine Learning would aim to automate this decision-making process by gathering loads of past data about various individuals and their ability to pay back loans within the life of the loan and deduce patterns and relationships between financial behavior and loan defaulting.

But what are models? Models are mathematical abstractions about the data we deal with. Most models have their roots in classical statistics and rely on quantitative analysis of the data to gauge relationships in it. Models encapsulate the decision-making process that ties a row of data to a label that describes it.

Once we have our data and our model, we ‘train’ the model on the data and then being to throw new instances at it and see how it performs – evaluation. This last step is crucial for any machine learning method, since we would be wasting our time if our algorithm was no better than random guessing.

This first chapter will involve learning how to build your first classification model from scratch, and then how to apply an evaluation method to see how well it does.