THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

MODULE 1 UNIT 1
Lesson Infographic 1 Transcript

IN COLLABORATION WITH getsmarter™

Finding the most suitable machine learning technique

Introduction

This is a roadmap to matching the most appropriate machine learning technique with a given problem. First, you need to determine if you are working with labelled data or unlabelled data.

1. Labelled data

This option is chosen if the data to be used for the machine learning technique is labelled. If the data is labelled, supervised machine learning is applicable.

1.1 Supervised learning

From there, a further choice must be made between predicting values or predicting categories, depending on the goal of your machine learning application.

1.1.1 Predicting values

When predicting values, regression is the most appropriate machine learning method to apply. Regression can include *k*-nearest neighbour regression, linear regression, Bayesian linear regression, neural network regression, random forests, or boosted decision tree regression.

1.1.2 Predicting categories

When predicting categories, two-class classification is the most appropriate method to use. This can include two-class logistic regression, two-class averaged perceptron, Bayes classifier, two-class decision forest, two-class boosted decision tree, or two-class neural network.

2. Unlabelled data

This option is chosen if the data in question is unlabelled. If the data is unlabelled, unsupervised machine learning is applicable.

2.1 Unsupervised machine learning

From there, a further choice must be made between assembling features or discovering structure in data, depending on the goal of your machine learning application.

2.1.1 Assembling features

When assembling features, clustering is the most appropriate machine learning technique to apply, like *k*-means clustering.

2.1.2 Discovering structure

If the goal is to discover structure in data, dimension reduction is the best machine learning method to apply, like principal component analysis (PCA).

Conclusion

This is a helpful tool to follow when presented with data and tasked with matching an appropriate machine learning technique to solve a problem.