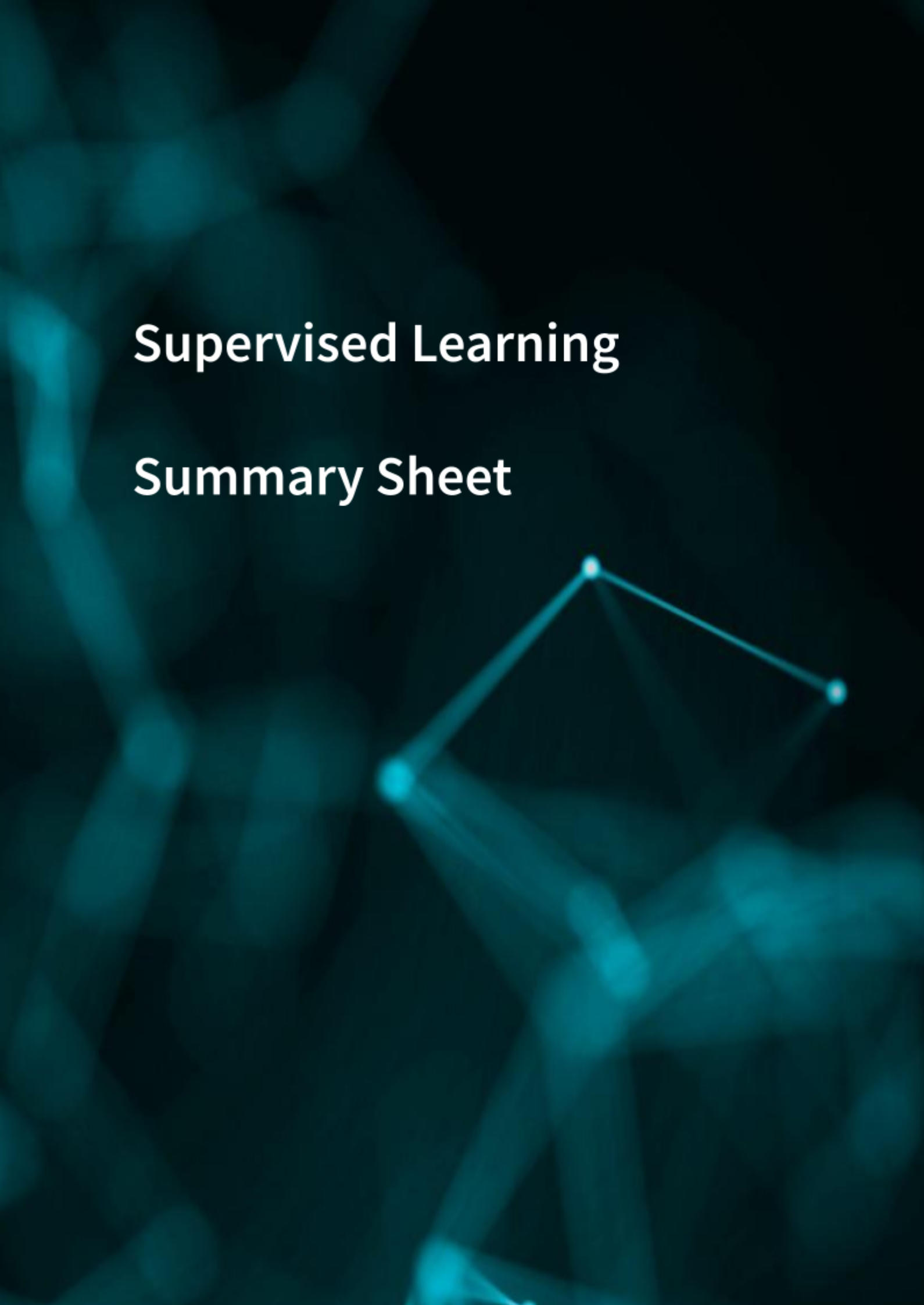
Supervised Learning

Summary Sheet



Supervised Learning Summary Sheet

Learning objectives - Through this module you will learn....

- About basic discriminative classifiers such as K-Nearest Neighbours and basic generative classifiers such as Naive Bayes.
- About formulation of a machine learning problem as a regression or classification task.
- How to use the right metrics and evaluation process to estimate the performance of a model.
- How to use a range of techniques to tune your supervised learning model in order to improve performance.
- About Decision Trees, a simple but powerful machine learning model for classification.

Overview

This module covers the basics of supervised learning and equips you with the knowledge required to solve a supervised learning problem from end-to-end: from processing the data, to building the model, to assessing how well it's doing, and evaluating it on test data. You will learn about the differences between regression and classification problems and which models are suitable for the different cases.

Why this topic matters

Supervised learning is the most straightforward method of teaching machines how to learn. By collecting labelled data and setting up our models and training regimes correctly, we can have machines automatically learn relationships between inputs and desired outputs. It's a core concept that every data scientist or machine learning practitioner needs to understand. Understanding the basics of supervised learning will allow you to better understand a whole range of models, from ones as simple as linear regression all the way to complex ones like deep neural networks.

Practical use cases and applications

If you can think of a problem being solved in the real world today, there's a good chance it's a supervised learning problem. From facial recognition to predicting the weather, these are all examples of applications powered by supervised learning.

This article details amazing work by companies like IBM, Twitter, and Yelp. All these applications are examples of cutting edge supervised learning.

Another cool application using supervised learning is one you might have heard of, FaceID. By utilizing supervised learning, Apple is able to provide its users with easier and more secure access to their devices.

Supervised Learning Summary Sheet

How Supervised Learning creates value and can be useful to organisations

Supervised learning allows companies to automate a wide variety of tasks. Given a good, labelled dataset, an algorithm can learn the required patterns and then carry out predictions on new data. These predictions can generally be made in the fraction of the amount of time a human needs which allows for faster delivery of services. Additionally, once trained, the same algorithm can be deployed to a host of different machines, allowing a company to scale its services to serve a huge number of customers with minimal overhead.

Module Components - Time & Difficulty

LU/PAK	Est. Time to complete	Difficulty Level
L01 Classic Discriminative Models	2hr	Medium
L02 Naive Bayes	1hr	Medium
L03 Model Evaluation	1hr	Easy
L04 Model Validation	2hr	Medium
L05 Decision Trees	2hr	Medium
Assignment - Classification and Model Selection with Kickstarter	1hr	Medium
Assignment - Analysis of Asteroid	1hr	Medium
Assignment - Analysis of Heating Load	2hr	Hard
Assignment - Classification of counterfeit bank notes	2hr	Hard

Tools, technique, libraries covered in this module

- Numpy
- Pandas
- Scikit Learn
- Scipy

Prerequisite knowledge needed to be able to work with this module

- The Data Science Toolbox
- Maths for Data Science
- Introduction to Machine Learning

Supervised Learning Summary Sheet

Which topics we will cover, and why

We will cover the theory behind discriminative classifiers and generative classifiers. Getting a good grasp of this theory will help learners decide how to frame future problems and what category of classifier will work best. To solidify understanding and expand the learner's toolkit, we will also look at specific models that fall under these categories. We will examine K-Nearest Neighbours and Decision Trees which are examples of discriminative classifiers and Naive Bayes which is an example of a generative classifier. Aside from classifiers, we will also look at Linear Regression which is suitable for solving a range of regression problems. On top of learning about models, learners will also study the impact of data on their models and how to properly assess model performance. This is key for ensuring that deployed models perform well in the field.

Which other topics are NOT encompassed in this module

This module will not cover more advanced supervised learning models like deep neural networks or gaussian processes. We refer the learner to the suggested additional readings to learn about these models and note that deep neural networks are covered in the Neural Networks and Deep Learning module. This module will also not cover supervised learning for time series data but this is covered in Time Series Analysis.

Suggested additional readings

- Bengio, Y., Goodfellow, I. and Courville, A., 2017. Deep learning (Vol. 1). Massachusetts, USA:: MIT press.
- Rasmussen, C.E., 2003, February. Gaussian processes in machine learning. In Summer school on machine learning (pp. 63-71). Springer, Berlin, Heidelberg.

CAMBRIDGE SPARK

THE SPECIALIST AI & DATA SCIENCE LEARNING PARTNER FOR ORGANISATIONS

EDUKATE.AI