

P S DS ML AI

June 1, 2023

Contents

1	Machine Learning	5
1.1	Supervised Learning	6
1.1.1	Classification	6
1.2	Unsupervised Learning	6
1.3	Measuring Model Performance	6
1.3.1	Cross Validation	6

Chapter 1

Machine Learning

1.1 Supervised Learning

1.1.1 Classification

KNN

K-Nearest-Neighbours - The euclidean distance between a data point and a series of features is calculated and the majority value of the target variable is taken.

$$distance = \sqrt{(w_0 - w_1)^2 + (x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2 \dots}$$

- A large value of k leads to a less complex model and can lead to *underfitting*
- A smaller value of k increases the complexity of the model and can lead to *overfitting*

1.2 Unsupervised Learning

1.3 Measuring Model Performance

1.3.1 Cross Validation