Chapter 3. Feature Selection and Feature Engineering

Feature engineering is the first step in a machine learning pipeline and involves all the techniques adopted to clean existing datasets, increase their signal-noise ratio, and reduce their dimensionality. Most algorithms have strong assumptions about the input data, and their performances can be negatively affected when raw datasets are used. Moreover, the data is seldom isotropic; there are often features that determine the general behavior of a sample, while others that are correlated don't provide any additional pieces of information. So, it's important to have a clear view of a dataset and know the most common algorithms used to reduce the number of features or select only the best ones.

Previous Section (/book/big_data_and_business_intelligence/9781785889622/2/ch02lvl1s

Next Section (/book/big_data_and_business_intelligence/9781785889622/3/ch03lvl1sec2