Handson Machine Learning Notes

Part I

Supervised/UnSupervised/SemiSupervised/Reinforcement

Dimensionality Reduction/ Feature Extraction

Batch/Online Learning(train the system incrementally by feeding it data instances

sequentially)

Part II

Pipe line:A sequence of data processing components is called a data pipeline

strati‐ed sampling: the population is divided into homogeneous subgroups called strata, and the right number of instances is sampled from each stratum to guarantee that the test set is representative of the overall population

Replace NA value, median,delete,zero

Categorical-numerical(binary)

There are two common ways to get all attributes to have the same scale: min-max

scaling and standardization.

Cross-Validation