Cairo University

Applications of Machine Learning in Chemical Engineering

Tutorial 2: Review of ML Algorithms

Eng/ Samer Hany

Agenda

Review of ML Algorithms

- Quick Review of ML Algorithms
- Quick Review of OOP
- Introduction to PyTorch

QUICK REVIEW OF ML ALGORITHMS

Quick Review of ML Algorithms

LINEAR REGRESSION

- Learning Type:
  - Supervised
- Advantages:
  - Simple and doesn't require high computation power
  - Very easy to understand and interpret
- Disadvantages:
  - Limited to linear relations
  - Sensitive to noise and outliers
  - Prone to overfitting
  - Sensitive to multi-collinearity

Quick Review of ML Algorithms

LOGISTIC REGRESSION

- Learning Type:
  - Supervised
- Advantages:
  - Simple and doesn't require high computation power
  - Good performance with small datasets
  - Output can be interpreted as a probability
- Disadvantages:
  - Assumes linearity
  - Poor performance with highly correlated features
  - Built for binary classification by default (becomes more complicated for multi-class problems)

Quick Review of ML Algorithms

K-NEAREST NEIGHBORS (KNN)

- Learning Type:
  - Supervised
- Advantages:
  - Simple and intuitive to understand
  - Automatically adapts as new data is collected
  - Easy to implement for multi-class classification
  - Can be used for classification and regression
- Disadvantages:
  - Number of neighbors must be defined by the user
  - Requires relatively high computational power
  - Doesn't work well with imbalanced data
  - Sensitive to outliers

Quick Review of ML Algorithms

SUPPORT VECTOR MACHINES (SVM)

- Learning Type:
  - Supervised

- Advantages:
  - Works well with complex non-linear data
  - Relatively low risk of over-fitting
  - Works well with unstructured data
  - Can be used for classification and regression

- Disadvantages:
  - Kernel function must be provided by the user
  - Requires high computation power
  - Difficult to understand and interpret (with complex kernels)

Quick Review of ML Algorithms

DECISION TREES

- Learning Type:
  - Supervised
- Advantages:
  - Simple to understand and interpret
  - Able to handle both numerical and categorical data
  - Has built-in feature selection
  - Performs well with large number of features
  - Directly handles multi-class problems
- Disadvantages:
  - Highly Prone to overfitting especially with small datasets
  - Sensitive to data changes (tree structure can vary significantly depending on the training data)

Quick Review of ML Algorithms

RANDOM FOREST

- Learning Type:
  - Supervised
- Advantages:
  - Less prone to overfitting than decision trees
  - Good performance on imbalanced datasets
  - Good handling of missing data and outliers
  - Useful to extract feature importance
- Disadvantages:
  - Requires high computational power
  - More complex and less interpretable than decision trees

Quick Review of ML Algorithms

ARTIFICIAL NEURAL NETWORK (ANN)

- Learning Type:
  - Supervised & Unsupervised
- Advantages:
  - Can be used to model highly complex data
  - Can be used for classification and regression
  - Can be used for a variety of tasks (e.g. image recognition, natural language processing, time series prediction)
  - Can continuously improve with more data
- Disadvantages:
  - Difficult to interpret
  - There are many parameters to fine-tune (e.g. layers, nodes, etc)
  - Require high computational power
  - Prone to overfitting

Quick Review of ML Algorithms

K-MEANS CLUSTERING

- Learning Type:
  - Unsupervised
- Advantages:
  - Relatively easy to implement and understand
  - Scales well to large datasets
- Disadvantages:
  - Number of clusters must be defined by the user
  - Poor performance with clusters of different sizes, shapes, and densities.

Thank you

- Eng/ Samer Hany
  samer.hany@eng.cu.edu.eg
- Eng/ Nada Ashraf
  bakrynada8@eng.cu.edu.eg