Lesson 4.3

Supervised Learning: Classification

Two of Machine Learning's fundamental approaches: **supervised** and **unsupervised** learning.

supervised learning.
*In a **classification** problem, the outputs are categorical or discrete.*

Categories of Algorithms

Some of the most common types of classification problems include:

- *Classification on tabular data: The data is available in the form of rows and columns, potentially originating from a wide variety of data sources.*
- *Classification on image or sound data: The training data consists of images or sounds whose categories are already known.*
- *Classification on text data: The training data consists of texts whose categories are already known.*

Examples:
- Computer vision
- Speech recognition
- Biometric identification
- Document classification
- Sentiment analysis
- Credit scoring in banking
- Anomaly detection

Categories of Algorithms

At a high level there are three main categories of classification algorithms.

- **Two-class/Binary Classification**
  - The classifier chooses from only two categories; each output belongs to one or the other.

- **Multi-Class Single Label Classification**
  - The classifier chooses from multiple categories; each output belongs to a single category only.

- **Multi-Class Multi-Label Classification**
  - The classifier chooses from multiple categories; each output can belong to one or more categories.

Specific Algorithms for each Categories

Two-Class Classification Algorithms

Predict between **two** categories (binary classification)

Algorithm	Characteristics
Two-Class Support Vector Machine	Under 100 features, linear model
Two-Class Average Perception	Fast training times, linear model
Two-Class Decision Forest	Accurate, fast training times
Two-Class Logistic Regression	Fast training times, linear model
Two-Class Boosted Decision Tree	Accurate, fast training, large memory footprint
Two-Class Neural Network	Accurate, long training times

Multi-Class Classification Algorithms

Predict between **several** categories (binary classification)

Algorithm	Characteristics
Multi-Class Logistic Regression	Fast training times, linear model

Multi-Class Neural Network	Accurate, long training times
Multi-Class Decision Forest	Accurate, fast training times
Multi-Class Boosted Decision Tree	Non-parametric, fast training times and scalable
One-vs-All Multi-Class	Depends on the underlying two-class classifier

Lesson 4.7

Multi-Class Algorithms

Multi-Class Algorithm Hyperparameters

Multi-Class Logistic Regression
- Optimization tolerance
- Regularization weight

Multi-Class Neural Network
- Number of hidden nodes
- Learning rate
- Number of learning iterations

Multi-Class Decision Forest
- Resampling method
- Number of decision trees
- Maximum depth of the decision trees
- Number of random splits per node
- Minimum number of samples per leaf node

Note: the following charts or metrics can be used when evaluating results of a classification algorithm

- ROC curve
- Confusion matrix
- Recall

Notes: **Lesson 4.14**

Supervised Learning: Regression

*In a **regression** problem, the output is numerical or continuous.*

Introduction to Regression

Common types of regression problems include:

- *Regression on tabular data:* The data is available in the form of rows and columns, potentially originating from a wide variety of data sources.
- *Regression on image or sound data:* Training data consists of images/sounds whose numerical scores are already known. Several steps need to be performed during the preparation phase to transform images/sounds into numerical vectors accepted by the algorithms.
- *Regression on text data:* Training data consists of texts whose numerical scores are already known. Several steps need to be performed during the preparation phase to transform text into numerical vectors accepted by the algorithms.

Examples

- Housing prices
- Customer churn
- Customer Lifetime Value
- Forecasting(time series)
- Anomaly detection

Categories of Algorithms

Regression Algorithms

Linear Regression (Fast training, linear model)

- Linear relationship between **independent variables** and a **numeric outcome**
- Approaches
  - Ordinary least square method
  - Gradient descent

Decision Forest Regression(Accurate, fast training times)

- Ensemble learning method using multiple decision trees
- Each tree outputs a distribution as a prediction
- Aggregate to find a distribution closest to the combined distribution

Neural Network Regression(Accurate, long training times)

- Supervised learning method
- Label column must be a numerical data type
- Input layer + One hidden layer + Output layer

DESCRIPTION	HYPERPARAMETER
Learning rates	A value that defines the step taken at each iteration, before correction
L2 regulation weight	Penalize models to prevent overfitting
Number of learning iterations	The maximum number of times the algorithm processes the training cases.
Gradient Descent	A method that minimizes the amount of error at each step of the model training process

Lesson 4.18

Automate the Training of Regressors

Conventional machine learning process

- Features available in the data sets
- Algorithms that are suitable for the task
- Hyperparameters tuning
- Evaluation metrics

What is Automated Machine Learning?
Intelligently test **multiple** algorithms and hyper-parameters in parallel and return the best one

- Deploy into production
- Further customize or refine

Notes: Automated Machine Learning gives users the option to automatically scale and normalize input features. It also gives users the ability to enable additional featurization, such as missing values imputation, encoding, and transforms.

*The default behavior in Automated Machine Learning to deal with missing values for categorical features is to **Impute with most frequent value***

Lesson 4.21

Unsupervised Learning

*In **unsupervised learning**, algorithms learn from unlabeled data by looking for hidden structures in the data.*
The term unsupervised comes from the fact that, not having the expected outputs, the algorithm attempts to find on its own hidden structures in the data.

*Obtaining unlabeled data is comparatively inexpensive and unsupervised learning can be used to uncover very useful information in such data. For example, we can use **clustering** algorithms to discover implicit grouping within the data, or use **association** algorithms to discover hidden rules that are governing the data (e.g., people who buy product A also tend to buy product B).*

Types of Unsupervised Machine Learning

Clustering
Organizes entities from the input data into a finite number of subsets or clusters

Feature Learning(a.k.a Representation Learning)
Transforms sets of inputs into other inputs that are potentially more useful in solving a given problem

Anomaly detection
Identifies two major groups of entities

- Normal
- Abnormal(anomalies)

the following are examples of unsupervised learning algorithms -

- K-Means Clustering
- Principal Component Analysis (PCA)
- Autoencoders

Lesson 4.22

Semi-Supervised Learning

Sometimes fully labeled data cannot be obtained, or is too expensive—but at the same time, it may be possible to get partially labeled data. This is where **semi-supervised learning** is useful.

Semi-supervised learning *combines the supervised and unsupervised approaches; typically it involves having small amounts of labeled data and large amounts of unlabeled data.*

Approaches to Semi-Supervised Machine Learning

- **Self-training** - The model is trained with the labeled data, then used to make predictions for the unlabeled data (resulting in a dataset that is fully labeled).
- **Multi-view training** - Multiple models are trained on different views of the data (e.g., different feature selections, model architectures, etc.).
- **Self-ensemble training -** A Single model is trained on different views of the data (e.g., different feature selections, model architectures, etc.).

Lesson 4.24

Clustering

On this page, we'll discuss the unsupervised approach of *clustering* in more detail.

*As the name suggests, **clustering** is the problem of organizing entities from the input data into a finite number of subsets or clusters; the goal is to maximize both intra-cluster similarity and inter-cluster differences.*

Applications of Clustering Algorithms

- Personalization and target marketing
- Document classification
- Fraud detection
- Medical imaging
- City planning

Clustering Algorithms

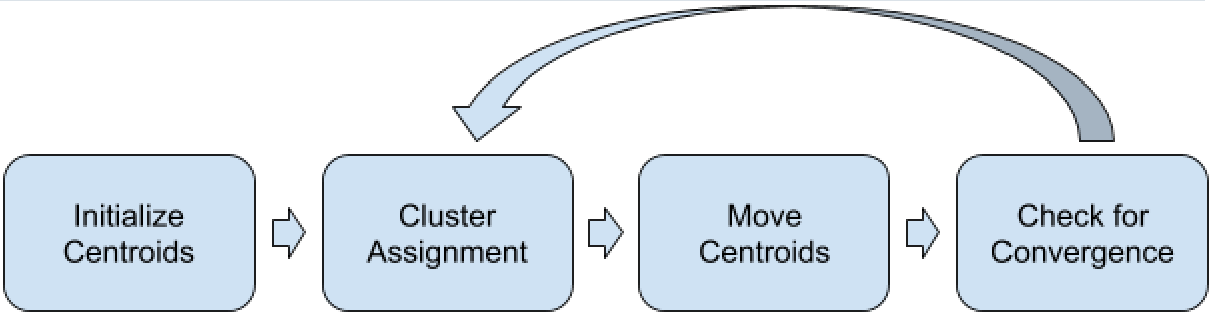
- **Centroid-based Clustering** - Groups members based on their distance from the center of the cluster.

- **Density-based Clustering** - Groups members based on how closely they are packed together; can learn clusters of arbitrary shape.
- **Distribution-based Clustering** - Groups members based on the probability of a member belonging to a particular distribution.
- **Hierarchical Clustering** - Builds a tree of clusters.

K-Means Clustering

K-means clustering algorithm

- K-Means is a centroid-based, unsupervised clustering algorithm.
- It creates up to a target (K) number of clusters and groups similar members together in a cluster.
- The objective is to minimize intra-cluster distances



K-Means Module Configurations

- Number of centroids
- Initialization approach
- Distance metric-Euclidean
- Normalize feature
- Assign label mode
- Number of Iterations