2024 Google ML Training Programme (Phase III)

Session 1.2

Types of Machine Learning and Algorithms

June 24, 2024

Godfrey A. Mills
Email: gmills@ug.edu.gh
Phone: 020-549-6944

Types of Machine Learning

- Types of learning rules or styles used in neural networks are:
  - **Supervised** (or predictive) learning >> task of learning a mapping function from an input to an output based on given labelled set of input-output pairs.
  - **Unsupervised** (or descriptive) learning >> learning task that discovers patterns or inferences from dataset that consists of input data without labeled responses or teacher.
  - **Semi-supervised** learning >> learning mechanism where the training data usually includes a few desired outputs
  - **Reinforcement**/hybrid learning >> learning task based on self (rewards for correct actions) to achieve given goals

Supervised learning protocol

- Supervised type of learning involves an external teacher.
  - In supervised learning we consider systems that applies a function **f(.)** to input items **x** to return an output **y = f(x)**
  - Item **x** (or instance) is drawn from an input space **X** and output **y** (or label) is drawn from an output space **Y**.
  - Output function **y = f(x)** is the target function and the output **y = g(x)** is the learned model
  - In supervised learning, we deal with systems or processes whose function f(**x**) is learned from examples.
  - We typically use machine learning when the function f(**x**) we want the system to apply is unknown to us.

Supervised learning protocol

- **Cont'd….**
  - Supervised learning process can be described by a feedback system where knowledge is represented by a set of labeled training data that is given to the learning algorithm to return a learned model g(x)
    - >> Dtrain = $\{x_1,y_1\}$, $\{x_2,y_2\}$, ……, $\{x_n,y_n\}$
- Learning is the adjustment of parameters of the model under the influence of training examples.
  - Model g(.) is applied with test samples to return predicted output labels which is compared with test labels
  - >> Dtest = $\{x^*_1,y^*_1\}$, $\{x^*_2,y^*_2\}$, ……, $\{x^*_n,y^*_n\}$

Supervised learning protocol

- **Cont'd….**
- Supervised learning techniques are very good for handling the following nature of problems or tasks :
  - Binary class classification >> dividing output response or data into two classes >> 2 possible labels, f(x) = {-1, +1}
  - Multiclass classification >> choosing between more than two types of outputs >> k possible labels, f(x) = {1, 2,,
  - Regression modelling >> predicting continuous values >> numerical (labels are continuous valued, linear/polynomial
  - Ensemble model >> combining the output predictions or response of multiple ML models (or base learners)

Supervised learning protocol

- **Cont'd….**
- Standard supervised learning algorithms:
  - Naïve Bayes (NB)
  - K-Nearest Neighbour (KNN)
  - Decision Tree (DT)
  - Support Vector Machine (SVM)
  - Random Forest (RF)
  - Linear Regression (LG)***

Unsupervised learning protocol

- Unsupervised learning (also called knowledge discovery) is a type of learning performed without any external guidance to discover hidden statistical patterns in data and cluster them.
  - the kinds of patterns to look for is not given.
  - there is no metric to use (unlike supervised learning where the predicted labels can be compared with observed output)
- In unsupervised learning, some form of self organization is done to produce desired outputs >> degree of similarity used
- Unsupervised learning problem types are of two forms:
  - >> clustering ;
  - >> association ;

Unsupervised learning protocol

- **Cont'd….**
- Unsupervised learning techniques are very good for handling the following nature of problems or tasks :
  - Clustering >> splitting dataset into groups based on the level or degree of similarity.
  - Anomaly detection >> identifying unusual data points in a data set.
  - Association mining >> identifying sets of items in a data set that frequently occur together.
  - Dimensionality reduction >> reducing the number of variables in a data set.

Unsupervised learning protocol

- **Cont'd….**
- Standard unsupervised learning algorithms often used are:
  - K-Means >> clustering
  - Principal Component Analysis (PCA) >> data dimension reduction
  - Self-Organizing Map (SOM) Neural Network >> clustering
  - Isolation Forest >> clustering

Reinforcement learning protocol

- This is a type of learning with limited guidance >> it lies between the supervised learning and unsupervised learning.
  - Learning algorithm is only provided with a grade/score as a critic or agent which dictates the network performance
  - Learning of an input-output mapping is performed through continuous interaction with environment with objective to minimize a cost function based on cumulative actions taken
  - Reinforcement learning consists of 3 primary components:
    - agents (learning agent)
    - environment (agent interacts with environment)
    - actions (agents can take actions).

Reinforcement learning

- **Cont'd….**
  - An agent learns from the environment by interacting with it and it receives ***rewards*** and ***penalty*** for performing actions
  - The agent selects and performs actions in a sequence of time steps and receives rewards for performing correctly and penalties for performing incorrectly.
  - Unlike unsupervised learning, the goal for reinforcement learning is to find a suitable action model or cost function that would lead to minimized ***total cumulative cost of actions taken by the agent*** over a sequence of time steps instead of the immediate cost >> i.e., ***maximize rewards of actions taken by the agent***

Reinforcement learning

- **Cont'd….**
- Reinforcement learning techniques are very good for handling the following nature of problems or tasks :
  - Robotics >> train robots to learn and perform tasks of the physical world.
  - Video gameplay >> training bots to learn how to play many video games.
  - Resource management >> given finite resources and a defined goal, the algorithm can help enterprises to plan out how to allocate resources.

Semi-Supervised learning

- **Semi-supervised** learning >> learning mechanism where the training data usually includes a few desired outputs
  - In this type of learning, a mixture of the two learning methods (supervised and unsupervised) is used.
  - Algorithm is usually fed with a small amount of labelled training dataset, and from this, the algorithm learns the dimensions of the data set which it then applies on new unlabeled data.
  - Algorithm is usually free to explore the data on its own and develop its own understanding of the data set.

Semi-Supervised learning

- **Cont'd….**
- Semi-supervised learning techniques are very good for handling the following nature of problems or tasks :
  - Machine translation >> training algorithms to translate language based on less than a full dictionary words.
  - Fraud detection >> identifying cases of fraud when only few positive examples are available.
  - Labelling data >> training algorithms on small data sets to learn and apply the data labels created to larger sets automatically.

Ensemble learning

- **Ensemble** learning >> a composite learning mechanism that combines several machine learning algorithms into a single predictive model to create improved performance.
  - In this type of learning, individual machine learning algorithm votes and the final prediction label returned is one that performs majority voting.
  - All machine learning algorithms in ensemble are trained separately with same training and testing dataset.

Ensemble learning

- **Cont'd….**
- On basis of type of base learners, ensemble learning can be divided into two groups:
  - Homogeneous ensemble method >> uses the same type of learner in each iteration >> all supervised or all unsupervised learning models.
  - Heterogeneous ensemble method >> uses different type of base learner in each iteration >> hybrid of supervised and unsupervised learning models.

Ensemble learning

- **Cont'd….**
- Commonly used ensemble learning methods:
  - Bagging >> learning technique that combines multiple learners in a way to reduce the variance estimates
    - Example, Random Forest trains M Decision Trees on different random subset of data and perform voting for final prediction.
  - Boosting >> learning technique that uses a set of low accuracy (weak) classifier to create high accuracy (strong) classifier >> algorithm tracks the model that fails the accurate prediction.

Ensemble learning

- **Cont'd….**
- Three popular boosting algorithms used in ML model developments:
  - AdaBoost (adaptive boosting)
  - Gradient Tree Boosting
  - XGBoost
- **Stacking** >> learning technique that combines multiple base classification models predictions into a new data set >> this new data are treated as input data for another classifier which is employed to solve problem.

ML Tasks for Model Development

- Machine learning can be used to perform different tasks, but the most common tasks performed by machine learning are :
  - ***Classification*** algorithms >> problem of predicting which class (predict discrete values) a new observation should be placed, based on a training set of instances or observations.
    - ***2-class*** system (binary classification) >> where the class labels can be linearly separated into two disjointed classes.
    - ***K-class*** system (multiclass classification) >> class labels are not mutually exclusive and give multiple binary class.
    - Classification algorithms for classification tasks:
      - Naïve Bayes, Logistic regression, KNN, DT, RF, SVM.

ML Tasks for Model Development

- **Cont'd….**
  - ***Regression*** tasks >> problem of predicting continuous real values or labels given a new observation and a function that is based on a training set of instances or observations.
  - Unlike classification, regression involves finding correlations between dependent and independent variables.
    - Regression tries to find the best fit line that can predict the output more accurately, while classification finds the decision boundary that can divide the dataset into classes.
  - Why use regression and not classification ?? and vice versa
  - How different is regression and from classification?