ML_Assignment

January 19, 2022

1 Machine Learning (Lecture_1)

- Machine learning types
- Algorithms used in each type

2 What is machine learning

We can simply define that 'give training to a machine for a specific task and after training, it will be able to give prdictions on unseen related data smartly'. ## Types of machine learning - Supervised - Un-supervised/Clustering - Semi-supervised - Reinforcement

3 Supervised machine learning

In this type, we have dataset with annonations (labels/masks) and we simply divide dataset in parts (training/testing). Most of the time 80% data is given to the machine for learning and 20% for the testing. After that, we analyse performance and take predictions on unseen dataset and make decision on results. ## Algorithms for supervised ML - Logistic regression - K-neasrest Neighbors (K-NN) - Support Vector Machine (SVM) - Kernel SVM - Naive Bayes - Desision tree classification - Random forest classification

3.0.1 Logistic regression

Logistic regression used to predict the probability of a target variable. The nature of target or dependent variable would be only two possible classes (0 or 1, Yes or No, True or False) having binary target variables, but there can be two more categories of target variables that can be predicted by it. Based on those number of categories, Logistic regression can be divided into following types: - Binary or Binomial (depended variable has Two possible types) - Multinomial (depended variable has three or more possible un-ordered types e.g. (Type-A, Type-B, Type-C)) - Ordinal (depended variable has three or more possible ordered type e.g. (good, very good, best)) ### K-NN K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm and is mainly used for classification predictive problems in industry. - Lazy learning algorithm: lazy learning (uses all the data for training while classification) - Non-parametric learning algorithm: a non-parametric learning (doesn't assume anything about the underlying data)

3.0.2 SVM

Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms which are used both for classification. They are extremely popular because of their ability to handle

multiple continuous and categorical variables. The main goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH) and it can be done in the following two steps —

- First, SVM will generate hyperplanes (x-axis, y-axis) iteratively that segregates the classes in best way.

- Then, it will choose the hyperplane that separates the classes correctly (margin line).

3.0.3 Kernel SVM

SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. - linear - Non-linear - Polynomial - Radial basis function (RBF) - Sigmoid

3.0.4 Naïve Bayes

Naïve Bayes algorithms is a classification technique based on applying Bayes' theorem with a strong assumption that all the predictors are independent to each other. Naïve Bayes classification algorithm can be used for binary as well as multi-class classification problems both. - Types of Naïve Bayes - Gaussian Naïve Bayes (Gaussian distribution) - Multinomial Naïve Bayes (Multinomial distributio) - Bernoulli Naïve Bayes (binary (0s and 1s))

3.0.5 Desision tree classification

Decision trees can be constructed by an algorithmic approach that can split the dataset in different ways based on different conditions. Decisions tress are the most powerful algorithms that falls under the category of supervised algorithms. - Classification decision trees (categorical) - Regression decision trees (continous data) The two main entities of a tree are decision nodes, where the data is split. And leaves, where we got outcome. The questions are decision nodes and final outcomes are leaves.

3.0.6 Random forest classification

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. Random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

4 Un-supervised/Clustering

In this type, dataset has no information of annotations and machine has to be learn by itself. After drawing a reasonable pattern, it can give the predictions on the unseen dataset as well. ## Algorithms - K-means clustering - Hierarchical clustering - Probabilistic clustering

4.0.1 K-means clustering

K-means clustering algorithm computes the centroids and iterates until we it finds optimal centroid. It assumes that the number of clusters are already known. It is also called flat clustering algorithm. The number of clusters identified from data by algorithm is represented by 'K' in K-means. K-means follows Expectation-Maximization approach to solve the problem. The Expectation-step is used for assigning the data points to the closest cluster and the Maximization-step is used for computing the centroid of each cluster.

4.0.2 Hierarchical clustering

Hierarchical clustering is another unsupervised learning algorithm that is used to group together the unlabeled data points having similar characteristics. Hierarchical clustering algorithms falls into following two categories: - Agglomerative hierarchical algorithms (each data point is treated as a single cluster) - Divisive hierarchical algorithms (all the data points are treated as one big cluster)

4.0.3 Probabilistic clustering

In machine learning, a probabilistic classifier is a classifier that is able to predict, given an observation of an input, a probability distribution over a set of classes, rather than only outputting the most likely class that the observation should belong to.

5 Semi-Supervised

In this type, we have dataset which is not fully annonated, or we can say some data has labels information and some has missing information. So we can try or mix both supervised and unsupervised approaches to solve such problems.

6 Reinforcement

This one is advance type and it is hit & trial based learning of the machine. Basically, it learns from the mistakes and as we give and guide the machine about its performance, it learns and find a suitable pattern for our dataset.Reinforcement learning is all about learning from the environment through interactions.For example, initially, the baby can't walk so you support the baby as much as you can, but you can't be there all the time so the baby has to learn it by itself. You can't teach the baby fully supervised, cause we know many things fall in the process. so the baby has to learn it from experience (*trial and fail*). ## Algorithms - Model-Free reinforcement learning - Policy optimization - Q- Learning - Model based reinforcement learning - Learn the model - Given the model

6.0.1 Model-Free reinforcement learning

The model stands for the simulation of the dynamics of the environment. It predicts what the environment will give next. Model-free algorithms rely on trial-and-error to update its knowledge. As a result, it does not require space to store all the combination of states and actions. Policy optimization and Q-Learning are examples of model-free learning algorithm. - Policy optimization: You can think of policy as a strategy or agent's behavior function. Finding the optimal policy / optimal value functions is the key for solving reinforcement learning problems. The optimal policy

is the policy which maximizes the expected reward for each state. - Q-learning: Off policy learning method , means, it uses the target policy to choose the best next action while following the behavior policy.

6.0.2   Model based reinforcement learning

Model-based algorithms have been shown to provide more sample efficient and generalizable learning. In model-based deep reinforcement learning, a neural network learns a dynamics model, which predicts the feature values in the next state of the environment, and possibly the associated reward, given the current state and action. This dynamics model can then be used to simulate experiences, reducing the need to interact with the real environment whenever a new task has to be learned.

7   References

- RL
- Wiki
- Deep Math
- Model RL
- Tutorialspoint

[ ]: