AMRITA -CS-STAR

S3 TAG Activity 3.x

Workshops on ML Algorithms

"The best way to learn is to teach"

As part of **Collaborative learning** of machine learning algorithms, we have the following activity for the TAG teams

We are **assigning one machine learning algorithm for each group**, You are supposed to have discussions within the team , divide the tasks, have brain storming sessions within the channel to understand the concepts, select a suitable data set, do implementation of the algorithm and be ready to conduct a workshop to every other team. The workshop should have two phases- 1) **Presentation** which explains the concepts in depth.2) **Hands-on session** where you will make all the teams do the implementation of your ML algorithm.

Topic	Team
Linear regression	Group F- Durga
Logistic Regression	Group J - Abhishek
K- Nearest Neighbour	Group I- Sanjana
Naïve Bayes	Group D- Khyathi
Support Vector Machines	Group H- Harsha Vardhan
Decision Tree	Group A- Gayathri Vijay
Random Forest	Group G- Nikhil
Ensemble Techniques	Group C- Vinayak
Clustering	Group E- Shyam Prasad

All of you learn and practice 3.1a (data Exploration- Python pandas) and 3.1b (Steps in building a good machine learning model) before you get into the content of ML algorithms

7 Step model for building an efficient ML model (ref: Guo's framework)

1. Data collection
   → Defining the problem and assembling a dataset
2. Data preparation
   → Preparing your data

3. Choose model
4. Train model
   → Developing a model that does better than a baseline
5. Evaluate model
   → Choosing a measure of success
   → Deciding on an evaluation protocol
   →This is with validation data (testing). This may result in a feedback mechanism where you may end up changing the model as it doesn't give expected good result or you do hyper parameter tunings and again check the accuracy
   → The performance Evaluation Metrics will come to help here
6. Parameter tuning
   → Scaling up: developing a model that overfits
   → Regularizing your model and tuning your parameters
7. Predict (Testing : with real test data)

Action Plan (Activity 3.x)

Step I : Collaborative Concept Learning: (Steps to be followed by each team)

- Refer 3.1a Data Exploration using pandas- Python and get familiar with the techniques
- Refer 3.1b -Steps to build an efficient Machine learning algorithm, Get familiar with the terms hyper parameter tuning, cross validation, data preparation, over fitting, under fitting, Bias ,Variance, How to Building a model, Testing the model, Train Error, Test Error, Performance evaluation metrics etc
- Watch the videos and the links given and thoroughly understand the concept, algorithms, mathematics behind the algorithm assigned to you.
- Think through- What kind of data is this model suitable for? For what kind of problem will this method be suitable? What are the pros and cons of the method?
- Think through- What are the hyperparameters for your algorithm? What tuning can improve the performance of the algorithm?
- Think through-What performance Evaluation metrics fits here?
- Then you select a suitable data set to work on (refer the example code given/ or UCL repository/ Kaggle. To start with take a simpler data set)
- Do the data preparation/ pre-processing/ Data Cleaning (like Dealing with missing value/duplicates, Normalise the data in to range (0-1) or any other cleanup relevant to your data) to make it suitable to give for training
- Refer to the links given (source code Jupyter notebooks) and implement that ML algorithm- Training and testing (from input thru modelling thru validation thru performance evaluation) using python pandas inbuilt functions in scikit . Scikit provides the libraries for most of the above tasks.
    - Do cross validation/Split algorithm to train and test (70:30 or 80:20)

    - Train the model with the train data and get the model (fitting)
    - How many epochs/steps/batch size etc of training (whatever relevant to you)?
    - What is the convergence criteria of your training algorithm?
    - What is the train error you got? Does it fall into a case of over fitting or underfitting or good fit?
    - Do the testing of the algorithm with test set by passing the trained model
    - Do hyper parameter tuning (if required) play around and see how it improves the performance
    - Do a performance evaluation on the test results (refer the metric used for performance evaluation, generate graphs/tables, eg: Accuracy, Confusion matrix, ROC curve, Precision, recall etc for classification problem)
- Get "from the scratch" code for your algorithm from internet (Scikit site may have). Try to understand the code. Run the code with the data set. Do performance evaluation on the same.
- Write down the conclusions of your study

Step II : Workshop (Attendees : All other teams)

Based on the thorough understanding and implementation of the algorithm above you will conduct a workshop to your peer teams.

Things to note

- Prepare an impressive ppt (with illustrations/figures/visualisations/numerical examples if required) which explains the algorithm concepts/ Mathematics /step by step algorithm.
- Prepare a good tutorial material in jupyter notebook for each and every step of implementation with sufficient explanations Use the markdown feature in jupyter notebook: refer the link
  https://www.datacamp.com/community/tutorials/markdown-in-jupyter-notebook
  https://www.tutorialspoint.com/jupyter/jupyter_notebook_markdown_cells.htm
- Pass the Jupyter Notebook and Data set you worked on, to all the teams, one day before your workshop.
- The workshop should have sessions – which explains the concept and also handson sessions where all the teams implement the algorithm. Try to explain "from the scratch code" too of your algorithm to the best possible.

Step III: Post Workshop Tasks

- Update your githib repository by pushing this project and tutorial,,
- Update your resume with an entry of project done.

- Update linkedin profile
- Update your resume
- Are you interested to write a blog?, you can do it if interested. Write your story with the ML activity 3.x, your findings observations about the algorithm, Analysis on the performance of the algorithm on your data set etc.
  If you want to know how create a blog, refer the site below
  https://www.quora.com/What-is-blogging-and-how-can-I-start-my-own-blog

Put your best Efforts to give a good show. All the Best!!