CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Technology and Engineering

Department of Computer Science & Engineering

Date: 01 /12 /2022

Laboratory Manual

Academic Year	:	2022-23	Semester	:	6th
Course code	:	CS360	Course name	:	Machine Learning

Note: The laboratory will emphasize the use of Python, Python Packages, Machine Learning and its applications.

Instructions:

1. All Practical must be performed with different data sets from the list of datasets given.

Sr. No.	Aim	Hrs.	CO
	0Understanding Your World	2	1,2
1.	Numpy - Creating blank array, with predefined data, with pattern specific data - Slicing and Updating elements, - Shape manipulations - Looping over arrays. - Reading files in numpy - Use numpy vs list for matrix multiplication of 1000 X 1000 array and evaluate computing performance. For Help: https://www.dataquest.io/m/289-introduction-to-numpy/ https://cloudxlab.com/blog/numpy-pandas-introduction/ Pandas - Creating data frame - Reading files - Slicing manipulations - Exporting data to files - Columns and row manipulations with loops	4	1,2

	- Use pandas for masking data and reading if in Boolean format.
	For Help:
	https://www.hackerearth.com/practice/machine-learning/data-manipulation-visualisation-r-python/tutorial-data-manipulation-numpy-pandas-python/tutorial/
	Matplotlib
	    - Importing matplotlib
	    - Simple line chart
	    - Correlation chart
	    - Histogram
	    - Plotting of Multivariate data
	    - Plot Pi Chart
	For Help:
	https://towardsdatascience.com/data-visualization-using-matplotlib-16f1aae5ce70		

2.	Linear Regression
	Select the Dataset of your choice and respond to following questions.
	    - Why do you want to apply regression on selected dataset? Discuss the full story behind the dataset.
	    - How many total observations in data?
	    - How many independent variables?
	    - Which is dependent variable?
	    - Which are most useful variable in estimation? Prove using correlation.
	    - Implement linear regression using OLS method.
	    - Implement linear regression using Gradient Descent from scratch.
	    - Implement linear regression using sklearn API.
	    - Quantify goodness of your model and discuss steps taken for improvement (RMSE, SSE, R2Score).
	    - Discuss comparison of different methods.
	    - Prepare presentation for this work in group of 5
	For help: refer following free course on datacamp.
	Regression models: fitting them and evaluating their performance
		4	3,5

3.	Two Class Classification (Logistic Regression)
	Select Dataset of your choice and respond to following questions.
	    - Why you want to apply classification on selected dataset? Discuss full story behind dataset.
	    - How many total observations in data?
	    - How many independent variables?
	    - Which is dependent variable?
	    - Which are most useful variable in classification? Prove using correlation.
	    - Implement logistic function.
		4	3,5

	- Implement Log-loss function. - Implement Logistic regression from scratch. - Implement Logistic regression using sklearn API. - Quantify goodness of your model and discuss steps taken for improvement (Accuracy, Confusion matrices, F-measure). - Discuss comparison of different methods. - Prepare presentation for this work in group of 5 **For Help:** 1. https://medium.com/@anishsingh20/logistic-regression-in-python-423c8d32838b 2. https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python 3. **https://towardsdatascience.com/logistic-regression-python-7c451928efee** 4. https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8 https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html		
4.	Multi Class Classification (KNN) Select Dataset of your choice and respond to following questions. - Why you want to apply classification on selected dataset? Discuss full story behind dataset. - How many total observations in data? - How many independent variables? - Which is dependent variable? - Which are most useful variable in classification? Prove using correlation. - Implement KNN using sklearn api. - Implement code to find best value of k by splitting data in train and test. - Quantify goodness of your model and discuss steps taken for improvement. - Can we use KNN for regression also? Why / Why not? - Discuss drawbacks of algorithms such as KNN. - Prepare presentation for this work in group of 5 For Help: https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/	4	3,5
5.	Find a dataset with number of samples smaller than number of features. Apply principle component analysis to select K best features. Use Support Vector Machines/Naïve Bayes to train predictive model.	4	3,5

	Compare model accuracy and time required for training with full dataset and with selected K features. (use Sci-kit-learn library) https://scikitlearn.org/stable/auto_examples/applications/plot_face_recognition.html https://www.dataquest.io/blog/sci-kit-learn-tutorial/		
6.	Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.	2	3,5
7.	Practical Implementation of Principle Component Analysis(PCA). https://medium.com/machine-learning-researcher/dimensionality-reduction-pca-and-lda-6be91734f567	2	3,5
8.	Implement a Convolutional Neural Network (CNN) using Keras library. a) Implement a Convolutional Neural Network (CNN) for a handwritten Character Recognition. Use MNIST dataset to train the model. Generate test images by yourself. b) Case Study to build a CNN model using python. i) Build a dataset on home appliances (available at your home/ can take help from internet). Also use data augmentation technique to increase dataset. ii) Preprocess the image to fit into the model iii) Apply the CNN model and train over the preprocess data. iv) Evaluate the model using confusion matrix.	4	4,5
9.	Implement a RNN/LSTM to classify Text into categories according to the sentiment of the text. Make use of transfer learning by using pre-trained Word Vectors (Embeddings).	2	4,5
10.	Use K-Means Clustering and Hierarchical Clustering algorithm for following datasets.	2	4,5
11.	**Certification Course**		
12.	**Mini Project**	-	
	Total Hrs.	34	