#100daysOfMlCode - Day1

24 August 2018 08:43

Day 1	Machine Learning	#100daysofml

Humans have created huge amount of data in the complete human history but the pace of data creation now has increased enormously. After couple of years we will be sitting on humongous pile of data. Machine Learning is the potential to process that amount of data effectively.

Enviroment Setup

Anaconda -- spyder IDE for Python
RStudio for R

Part 1	Data Pre-processing

Figuring out the difference between dependent variables and independent variables.
In any ML model we use some independent variables to predict a dependent variable.

Importing the Libraries

Python Libraries : numpy, matplotlib, pandas
R Libraries : no libraries for now, may need later

Importing the Dataset

Python : pd.read_csv()

#100daysOfMlCode - Day2

25 August 2018 15:44

Day 2	Machine Learning	#100daysOfMlCode
Part 1	Data Pre-processing	

Python

Importing the Dataset
< pd.read_csv() >
Checking data format. Use the preferred number formatting.
Creating matrix of features/independent variables
< dataset.iloc[].values >
Creating dependent variable vector
< dataset.iloc[].values >
Missing Data - filling the missing data with the mean along the axis
< Imputer() > from sklearn.preprocessing

R

Importing the Dataset
< read.csv() >
! {indexes in R start from 1 unlike Python in which indexes start from 0}
Missing Data - filling the missing data with the mean along the axis
< ifelse(is.na()) >

Github -> https://github.com/amitforamit/sdsML

#100daysOfMlCode - Day3

26 August 2018 21:05

Day 3	Machine Learning	#100daysOfMlCode
Part 1	Data Pre-processing	

Since Machine Learning models are based on mathematical equations so we need to encode Categorical variables as numbers.

Python

Encoding Categorical Variable

< LabelEncoder() > from sklearn.preprocessing

Dummy Encoding

To avoid any order attribute in case of same level categorical variables, need to create dummy columns < OneHotEncoder() > from sklearn.preprocessing .It creates separate Columns for every categorical variables and assigning either 1 or 0 denoting the presence or absence of the variable.

R

Encoding Categorical Variable

Use < factor() > function without any order

Github -> https://github.com/amitforamit/sdsML

curquest