# Data-Science-Programming

## DSP-1 Covers various Topics

* Data set Preparation
* Types of Values
* Categorical Class, Nominal, Ordinal, Binary
* split of data (train & test)
* After getting the data ask these questions
  + 1. -What is the unit of analysis?
    2. -What are the predictors (features)?
    3. -What is the target/outcome variable?
    4. -Where is the data coming from?
    5. -Internal, external, combined?
    6. -How much data do you have (i.e., how many rows?)
    7. -What do to if data is too little or too much?
* -Discover and Visualize
* -Use visualization.
  + 1. -Histograms, boxplots, scatter plots
    2. -Use multiple variables in one plot when possible
* Observe descriptive statistics.
  + - 1. -How to interpret mean, median, std.dev, min, max?
* -Observe correlations
* -What is correlation?
* -Reduce variables (if needed) (how?)
* -Data Preparation/ feature engineering
* -Data Transformation
* -Train Models
* -Model Evaluation (Classification/Regression)

## DSP- 2 Regression

* How do we "train" the model: Find the set of betas that minimize RMSE (or, MSE – easier to do)
* 1) The Normal Equation (closed-form solution):
* X: Training set
* y: vector of output values
* Computationally very costly (if there are lots of features, or lots of data)
* 2) Optimizer: **Gradient Descent**
* Gradient Descent
* A generic optimization algorithm
* Tweak the parameters iteratively to minimize the cost function (SSE, MSE, RMSE, etc.)
* What do you mean by "minimize"???
* Lost in a mountain
* How do you get to the bottom? (i.e., "minimize")
* Feel slope below your feet!
* Then, go in the direction of steepest (descending) slope
* Diagram

  Description automatically generated with low confidence

Chart

Description automatically generated with low confidence

A picture containing diagram

Description automatically generated

Chart, line chart

Description automatically generated

* Stochastic Gradient Descent
* Mini-Batch Gradient Descent
* Polynomial Regression
* Regularization:
* L2 Regularization (Ridge Regression)
* L1 Regularization (Lasso Regression)
* Elastic Net
* Early Stopping
* Logistic Regression
* Softmax Regression

## DSP 3 : Support Vector Machines & Logistic Regression

## DSP 4: Decision Tree

## DSP 5: Ensemble Technicques

## DSP 6- Text Mining

## DSP 7- Neural Networks

## DSP 8- DNN- Deep Neural Networks (Keras)

Diagram

Description automatically generated

Diagram

Description automatically generated

## DSP 9- CNN – Convolutional Neural Nets (Keras)

A picture containing graphical user interface

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated

## DSP 10- RNN- Recurrent Neural Networks

Graphical user interface, text, application

Description automatically generated

Letter

Description automatically generated with low confidence

## DSP 11- Encoders

Graphical user interface, text, application

Description automatically generated

Chart, waterfall chart

Description automatically generated