Module 4: Supervised Classical Machine Learning

1 What is meant by **Supervised** ?

2 **Naive Bayes Classifier**

- 2.1 Bayes Theorem.
- 2.2 Continous Probability Distributions.
- 2.3 Estimation of population parameters for univariate case when data is continuous.
  - 2.3.1 Gradient Descent
  - 2.3.2 Newton's Method
- 2.4 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
- 2.5 **Coding Assignment 15.**
- 2.6 Biclass Bayes Classifier for Univariate case when the data is continuous.
- 2.7 Cross Validation and Testing
- 2.8 High Variance and High Bias problem.
- 2.9 **Portfolio Mini Project 1 on Bayes Classifier.**
- 2.10 Discrete Probability Distributions.
- 2.11 Estimation of population parameters for univariate case when the data is descrete.
- 2.12 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
- 2.13 **Coding Assignment 16.**
- 2.14 Biclass Bayes Classifier for Univariate case when the data is discrete.
- 2.15 Laplace Smoothing
- 2.16 **Portfolio Mini Project 2 on Bayes Classifier.**
- 2.17 Why the Bayes Classifier is called 'Naive' ?
- 2.18 Estimation of population parameters for multivariate case when the data is continuous and each variable has identical non stationary distribution.
- 2.19 Covariance Matrix Estimation, positive semi definiteness and singularity.
- 2.20 Correcting the problems with Covariance Matrix
  - 2.20.1 Fisher Method
  - 2.20.2 Diagonal Covariance Estimators
  - 2.20.3 Regularization Methods
- 2.21 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
- 2.22 **Coding Assignment 17.**
- 2.23 Biclass Bayes Classifier for Multivariate case when the data is continuous.
- 2.24 **Portfolio Mini Project 3 on Bayes Classifier.**
- 2.25 Estimation of population parameters for multivariate case when the data is descrete and each variable has identical non stationary distribution.
- 2.26 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
- 2.27 **Coding Assignment 18.**
- 2.28 Biclass Bayes Classifier for Multivariate case when the data is discrete.
- 2.29 **Portfolio Mini Project 4 on Bayes Classifier.**
- 2.30 Biclass Bayes Classifier for multivariate case when all the variables have different continuous and discrete probability distributions.
- 2.31 **Portfolio Mini Project 5 on Bayes Classifier.**
- 2.32 Multiclass Bayes Classifier for multivariate case when all the variables have different continuous and discrete probability distributions.
- 2.33 **Portfolio Project 4 on Bayes Classifier for Tabulated Data.**
- 2.34 **Portfolio Project 5 on Bayes Classifier for Visual Data (Images).**
- 2.35 **Portfolio Project 6 on Bayes Classifier for Textual Data (Text).**

3 **Linear and Polynomial Regression**
- 3.1 What is a Regressor ?
- 3.2 Dependent and Independent Random Variables.
- 3.3 Data Visualization
- 3.4 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
- 3.5 **Coding Assignment 19.**
- 3.6 Why and When Linear ?
- 3.7 High Bias V/s High Variance problem (Underfitting V/s Overfitting).
- 3.8 Population parameter estimation of Dependent Random Variable.
- 3.9 Likelihood Function.
- 3.10 Bayesian Inference.
- 3.11 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
- 3.12 **Coding Assignment 20.**
- 3.13 Linear Regression for Single Indepdendent Random Variable.
- 3.14 Improving Linear Regression.
  - 3.14.1 Normalization.
  - 3.14.2 Regularizarion.
    - 3.14.2.1 Classical Regularization Theory.
    - 3.14.2.2 Generalized Regularization Theory.
    - 3.14.2.3 Estimating the regularization parameter.
  - 3.14.3 Trade-off between stability and bias.
- 3.15 Different ways of looking at Regularization.
  - 3.15.1 Adding Prior Information about the practical world.
  - 3.15.2 Adding Penalty.

4 Cross Validation and Testing.

5 **Portfolio Mini Project 6 on Univariate Linear Regression.**

6 Natural Features and Derived Features.

7 Multivariate Linear Regression with numerous Natural Features.

8 **Portfolio Mini Project 7 on Multivariate Linear Regression.**

9 Multivariate Regression with Natural and Derived Features.

10 The interprettation.

11 Model selection for multivariate regression.
- 11.1 Kolmogorov Complexity Theory.
- 11.2 MDL princple.

12 Looking at Multivariate Regression through matrices.

13 Normal Equation.

14 Interprettation of Normal Equation of Multivariate Regression.

15 Regularization Revisited
- 15.1 Cross Correlation Matrices.
- 15.2 Positive Definite Cross Correlation Matrices.
- 15.3 Adding Prior information.
- 15.4 Adding Penalty.

16 **Portfolio Project 7 on Multivariate Regression.**

17 **Logistic Regression**
- 17.1 How come the word **Logistic** ?
- 17.2 Why Logistic ?
- 17.3 Dependent and Independent Random Variables.
- 17.4 Data Visualization
- 17.5 Dependent Random Variable as Binomial.
- 17.6 Exponential Family of Distributions.

41.5.1 Bagging.
41.5.1.1 Bagged Decision Trees.
41.5.1.2 Random Forest.
41.5.1.3 Extra Trees.

42 Boosting.

42.1 AdaBoost.
42.2 Stochastic Gradient Boosting.
42.3 XGBoost.
42.4 **Portfolio Project 9 on Boosting.**

43 **Time Series Forecasting**

43.1 What is Time Series Forecasting.
43.2 Time Series as Supervised Learning.
43.3 Load and Explore Time Series Data.
43.4 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
43.5 **Coding Assignment 24.**
43.6 Basic Feature Engineering.
43.7 Data Visualization.
43.8 Resampling and Interpolation.
43.9 Power Transforms.
43.10 Moving Average Smoothing.
43.11 Introduction to White Noise.
43.12 Introduction to Random Walk.
43.13 Stationarity in Time Series Data.
43.14 **Sample Python code demonstration along with detailed explanation for Coding Assignment.**
43.15 **Coding Assignment 25.**
43.16 Reframing Time Series Forecasting problems.
43.17 Seasonality and Trends.
43.18 Autoregression for Forecasting.
43.19 **Portfolio Mini Project 11.**
43.20 Moving Average Models for Forecasting.
43.21 ARIMA model for Forecasting.
43.22 **Portfolio Mini Project 12.**
43.23 Autocorrelation and partial Autocorrelation.
43.24 Grid Search for ARIMA Model Hyperparameters.
43.25 Forecast Confidence Intervals.
43.26 **Portfolio Project 10 on Time Series Forecasting.**

44 **Support Vector Machines**

44.1 Why Support Vector Machines ?
44.2 Linearly and Non linearly seperable datasets.
44.3 **Support Vector Machines Classification for Linearly Seperable Datasets or Hard Margin SVM.**
44.3.1 Optimal Seperating hyperplane.
44.3.2 No man's land.
44.3.3 Formulating the problem as maximizing the no man's land.
44.3.4 Constructing the contraints from the criteria of no man's land.
44.3.5 Non Convexity of Objective function.
44.3.6 Final Formulation of primal problem
44.3.6.1 Minimization of Primal Objective Function.
44.3.6.2 Convex Quadratic Programming.
44.3.6.3 Constrained Problem.