 **BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**WORK INTEGRATED LEARNING PROGRAMMES**

**COURSE HANDOUT**

**Part A: Content Design**

|  |  |
| --- | --- |
| **Course Title** | Regression |
| **Course No(s)** |  |
| **Credit Units** |  |
| **Course Author** | N.L.Bhanu Murthy |
| **Version No** |  |
| **Date** | 21/08/2018 |

**Course Description**

|  |
| --- |
| Regression is a type of supervised learning techniques wherein the target attribute is a continuous variable. This course focuses on developing a deeper understanding of regression models both from theoretical and implementation perspective. The model selection and performance measures will be discussed in this course. The issues with regression models like overfitting and the ways of combatting overfitting like ridge and lasso regression will be illustrated in this course. The interpretability/explicability of the models will also be discussed. |

**Course Objectives**

|  |  |
| --- | --- |
| **No** | Objective |
| **CO1** | Provide comprehensive algorithmic perspective of building regression models |
| **CO2** | Provide deeper understanding of overfitting and ways to combat overfitting |
| **CO3** | Provide competence to select appropriate model and performance measures |
| **CO4** | Provide hands-on to solve real life classification problems |
| **CO5** | Provide skill to interpret the predicted model |

**Text Book(s)**

|  |  |
| --- | --- |
| No | Author(s), Title, Edition, Publishing House |
| T1 | Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani: An Introduction to Statistical Learning, Springer |

**Reference Book(s) & other resources**

|  |  |
| --- | --- |
| No | Author(s), Title, Edition, Publishing House |
| R1 | Christopher Bishop: Pattern Recognition and Machine Learning, Springer International Edition |

**Content Structure**

<List down the modular content structure of the course either in the tabular form given below or as bullets>

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Title of the Module** | **Duration in mins** | **Reference** |
| M0 | Overview of certificate programme in ML & AI   1. Introduction to six modules of the programme 2. Programme Objectives & Learning outcomes 3. Evaluation of the courses (Quizzes/Assignments/Tests) 4. ML&AI in today’s world (by industry expert) 5. A real life ML&AI project and value of it to the business (by industry expert) | 20  15  15  25  25 | Class Notes |
| M1 | Introduction to Regression   1. Introduction to Supervised Learning 2. Regression vs. Classification 3. Linear and Polynomial Regression 4. Applications and Case Study for the module 5. Overview of Model Building for Linear Regression | 15  15  20  15  20 | T1 – Ch. 2 |
| M2 | Mathematics Foundations   1. First and Second derivatives of multivariate functions 2. Maxima and Minima of univariate and Multivariate Functions 3. Convex Function, Necessary and sufficient condition for convexity of functions 4. Determinant & Inverse of Matrices, Solving Simultaneous Equations | 20  15  20  15 | Class Notes |
| M3 | Model Building using Least squares   1. Cost/Loss Function for linear regression 2. Convexity of the Cost/Loss Function 3. Optimizing Cost/Loss Function by Solving Normal Equations 4. Implementation in Python 5. Optimizing Cost/Loss Function by Gradient Descent (I) 6. Optimizing Cost/Loss Function by Gradient Descent (II) 7. Optimizing Cost/Loss Function by Stochastic Gradient Descent and Batch Gradient Descent 8. Implementation in Python (Gradient & Stochastic Gradient Descent Methods) | 20  15  20  15  15  15  25  15 | R1 – Ch. 1  T1 – Ch. 2 |
| M4 | Model Accuracy & Selection   1. Measuring the Quality of Fit 2. Implementation in Python 3. Bias-Variance Decomposition 4. Training Data, Testing Data and Cross Validation Data 5. Polynomial Regression - Selecting the appropriate degree of the polynomial 6. Implementation in Python | 15  10  25  15  20  15 | T1 – Ch. 1  R1 – Ch. 1 |
| M5 | Overfitting   1. Introduction to Overfitting 2. Reasons for overfitting 3. Counters to control overfitting – Ridge Regression 4. Implementation in Python (Ridge) 5. Counters to control overfitting – Lasso Regression 6. Implementation in Python (Lasso) 7. Compare Ridge vs Lasso vs Model without Regularization with a case study | 20  15  20  10  15  10  15 | R1 – Ch. 1 |
| M6 | Interpretability of regression models   1. Statistics Foundations – Inferential Statistics and Hypothesis Testing, Significance tests, p-values (1) 2. Statistics Foundations – Inferential Statistics and Hypothesis Testing, Significance tests, p-values (2) 3. Interpretability of regression model through coefficients of the model 4. Interpretability of the regression built for the Case Study 5. Industry expert talk on regression for a real life business scenario | 20  20  15  15  30 | T1 – Ch. 3 |
|  |  |  |  |
|  |  |  |  |

**Weekly coverage of the course**

|  |  |
| --- | --- |
| **Week** | **Content / Assignments / Exercises** |
| Week1 | Video Content: M0, M1  Evaluative Quiz: Nil  Exercises (180 mins):   1. Identifying a given scenario is a regression or classification problem 2. Provide a real life business problem for regression and classification 3. Figure out the computational difficulties in solving linear regression and polynomial regression problems?   Assignment (120 mins)::   1. Python Program to read data from flat file and manipulate data with basic data operations   Minor Projects: Nil |
| Week 2 | Video Content: M2.1 to M2.4, M3.1 and M3.2  Evaluative Quiz: Nil  Exercises (180 mins):   1. Find the Del and Hessian of the given multivariate functions 2. Find the maxima/minima of the given multivariate functions 3. Check whether the following multivariate functions are convex or not. 4. For the given training data of 10 instances, show that the cost/loss function is convex.   Assignment (120 mins)::   1. Python Program to find determinant and inverse of a Matrix and thus solve system of 10 simultaneous equations in 10 variables   Minor Projects: Nil |
| Week 3 | Video Content: M3.3 to M3.8, M4.1 and M4.2  Evaluative Quiz: Q1 (preparation time – 300 mins)  Exercises/Assignments: Nil  Assignments: Nil  Minor Projects: Nil |
| Week 4 | Video Content: M4.3 to M4.6.  Evaluative Quiz: Nil  Exercises/Assignments: Nil  Minor Project 1 (300 mins)::  The case study (taught in M1.3.) should be implemented in Python using   1. Gradient Descent 2. Batch Gradient Descent 3. Stochastic Gradient Descent 4. Solving Normal Equation 5. To figure out the appropriate degree of the polynomial.   The project report should include comparative study of techniques outlined in (a), (b), (c), (d). It should also include comparative study with polynomial of different degrees (e). |
| Week 5 | Video Content: M5.1 to M5.7.  Evaluative Quiz: Q2 (preparation time – 300 mins)  Exercises/Assignment: Nil  Minor Projects: Nil |
| Week 6 | Video Content: M6.1 to M6.5.  Evaluative Quiz: Nil  Exercises/Assignments: Nil  Minor Project 2 (300 mins):  The case study (taught in M1.3.) should be implemented in Python to appreciate overfitting and implement ridge/lasso regression. The project report should include comparative study of ridge, lasso regression along with regression with no regularization. The project should also focus on interpretability and explicability of regression models and their connection to lasso regression. |
|  |  |

**Evaluation**

|  |  |  |
| --- | --- | --- |
| Evaluation Component | Marks | Type |
| Comprehensive Examination | 40% | Closed |
| Quizzes (2) | 24% | Open |
| 2 Minor Projects (Evaluated twice) | 24% | Open |
| Assignments/Exercises (2) | 12% | Open |

**Learning Outcomes**

|  |  |
| --- | --- |
| No | Learning Outcomes |
| LO1 | Ability to build appropriate regression model for a given real life business problem |
| L02 | Demonstrate the capability to select suitable degree of the polynomial regression and performance measures |
| LO3 | Ability to suggest appropriate methods to combat overfitting |
| LO4 | Ability to interpret the regression model |