 **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** | Dr. N.L.Bhanu Murthy |
| **Lead Instructor** | Dr. Chetana Gavankar |
| **Version No** | 3.0  (Modified by Dr. Chetana Gavankar with inputs from Dr. N.L.Bhanu Murthy) |
| **Date** | 23-March-2020 |

**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 regression 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**

|  |  |  |
| --- | --- | --- |
| **No** | **Title of the Module** | **Reference** |
| M1 | **Module 1: Introduction to Regression**   1. Introduction to Supervised Learning 2. Introduction to Regression and Classification 3. Linear and Polynomial Regression 4. [Error Function for Linear Regression](https://bits-pilani.instructure.com/courses/189/modules/items/10741) 5. [Introduction to Matrix Theory](https://bits-pilani.instructure.com/courses/189/modules/items/10740) 6. [Solving Simultaneous Equation with Matrices](https://bits-pilani.instructure.com/courses/189/modules/items/10739) | T1 – Ch. 2 |
| M2 | **Module 2: Building Simple Linear Regression Models**   1. [Mathematical Foundations 1 – Maxima and minima of function of one variable](https://bits-pilani.instructure.com/courses/189/modules/items/11073) 2. [Mathematical Foundations 2 - rank, Eigen values and eigen vectors, positive & negative definite/semi- definite matrices](https://bits-pilani.instructure.com/courses/189/modules/items/11080) 3. [Mathematical Foundations 3 - Maxima and minima of function of several variables](https://bits-pilani.instructure.com/courses/189/modules/items/11079) 4. [Convexity of errors function](https://bits-pilani.instructure.com/courses/189/modules/items/11078) 5. Building simple linear regression model by solving normal equations 6. Gradient descent algorithm 7. Gradient descent algorithm for a simplistic case 8. [Gradient descent, stochastic & mini-batch gradient descent algorithms](https://bits-pilani.instructure.com/courses/189/modules/items/11074) | Class Notes |
| M3 | **Module 3: Assessing Accuracy of Simple Linear Regression Models**   1. [Probability Foundations 1 – Discrete probability distributions](https://bits-pilani.instructure.com/courses/189/modules/items/11114) 2. [Probability Foundations 2 - Continuous probability distributions, normal distribution and t distributions](https://bits-pilani.instructure.com/courses/189/modules/items/11119) 3. [Accuracy of the coefficient estimates of the simple linear regression models](https://bits-pilani.instructure.com/courses/189/modules/items/11118) 4. Dependency of the dependent variable (target) variable on the independent variable (feature) 5. Accuracy of the simple regression model – RSE 6. Accuracy of the simple regression model – R^2 | R1 – Ch. 1  T1 – Ch. 2 |
| M4 | **Module 4: Building Multiple Linear Regression Models**   1. [Building multiple linear regression model by solving normal equations](https://bits-pilani.instructure.com/courses/189/modules/items/11129) 2. [Building multiple linear regression model by gradient descent algorithms](https://bits-pilani.instructure.com/courses/189/modules/items/11130) 3. [Performance measure for multiple linear regression models](https://bits-pilani.instructure.com/courses/189/modules/items/11133) 4. [Feature selection algorithms for multiple linear regression models](https://bits-pilani.instructure.com/courses/189/modules/items/11132) 5. [Forward and backward feature selection algorithms for multiple linear regression models](https://bits-pilani.instructure.com/courses/189/modules/items/11131) | T1 – Ch. 2 |
| M5 | **Module 5: Building Polynomial Regression Models, Overfitting and ways to combat overfitting**   1. [Polynomial Curve Fitting – Finding suitable degree of the polynomial for the problem](https://bits-pilani.instructure.com/courses/189/modules/items/11278) 2. Overfitting 3. [Ridge Regression and other ways of combatting overfitting](https://bits-pilani.instructure.com/courses/189/modules/items/11282) 4. [Lasso Regression](https://bits-pilani.instructure.com/courses/189/modules/items/11281) 5. [Bias-Variance Decomposition 1](https://bits-pilani.instructure.com/courses/189/modules/items/11280) 6. [Bias-Variance Decomposition 2](https://bits-pilani.instructure.com/courses/189/modules/items/11279) | R1 – Ch. 1 |

**Weekly coverage of the course**

|  |  |
| --- | --- |
| **Week** | **Content / Assignments / Exercises** |
| Week1 | Video Content: M1  Evaluative Quiz: Nil  Assignment: 1 week implementation time  Minor Projects: Nil |
| Week 2 | Video Content: M2  Evaluative Quiz: Q1 (based on M1 and M2 content)  Minor Projects: Nil |
| Week 3 | Video Content: M3  Minor Project 1 (300 mins):: 2 weeks implementation time  The case study should be implemented in Python |
| Week 4 | Video Content: M4.  Evaluative Quiz: Nil  Exercises/Assignments: Nil  Minor Projects: Nil |
| Week 5 | Video Content: M5  Evaluative Quiz: Q2 ((based on M3,M4 and and M5 content))  Exercises/Assignment: Nil  Minor Projects: Nil |

**Evaluation**

|  |  |  |
| --- | --- | --- |
| Evaluation Component | Marks | Type |
| Comprehensive Examination | 40% | Closed |
| Quizzes (2) | 24% | Open |
| Minor Project | 24% | Open |
| Assignment | 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 |