Birla Institute of Technology & Science, Pilani
Work-Integrated Learning Programmes Division
First Semester 2019-2020
Comprehensive Examination (Regular)

Course No.	: PCAM* ZC211
Course Title	: REGRESSION
Nature of Exam	: Closed Book
Weightage	: 40%
Duration	: 3 Hours
Date of Exam	: **02/11/2019 (FN)**

No. of Pages = 1
No. of Questions = 7

Q1. Suppose you asked to build a uni-variate regression model with 'N' training points. You are asked to fit a polynomial of degree 2 by minimizing the sum of squares of errors of training data points. [3 + 3 + 4 Marks]

(a) Show that the error function is convex function.
(b) Find out the exact (not an approximate) regression model that minimizes error the least. Hint: Gradient methods might bot help you here!
(c) If you are asked fit to a polynomial of degree 100, then list out the practical issues that would be faced in the above procedure.

Q2. Build the following linear regression models for the data set given below [4 Marks]

(a) $y = w_0$
(b) $y = w_0 + w_1 x$

Data Set:

x	-1	0	2
y	1	-1	1

Q3. Write down the steps to find R^2 value in single variate linear regression. Find R2 value for for the problem in the above question i.e., Q2. [4 Marks]

Q4. Can R^2 value be 1? If so, provide an example for which R^2 is 1. [4 Marks].

Q5. What are the two techniques to implement regularization for polynomial fitting? What is the difference between these two techniques? Explain the two techniques with all mathematical rigor. [6 Marks]

Q6. Do you agree that forward or backward stepwise selection algorithm guarantees the best optimal solution? If so, prove it? Otherwise what the issues are there in figuring out the best feature subset? [6 Marks]

Q7. Suppose you are a machine learning consultant and are given census data of 1,00,00,000 people containing 20 features to predict mortality. After doing few basic experiments, you decided to go ahead with regression. Discuss how you finalize the degree of the polynomial that you will be fitting for linear regression (whether you will be fitting linear, quadratic, cubic, curve to model the data). [6 Marks]