**Birla Institute of Technology & Science, Pilani**

**Work-Integrated Learning Programmes Division**

**PGP AIML**

Course No. :

Course Title : Regression

Nature of Exam : Closed Book

No. of Pages =**2**

No. of Questions = **4**

Weightage : 40%

Duration : 2 Hours

Date of Exam :

**Q.1.A).**Validate the statement “There is no difference between regression and classification except the format of the data and its labels”. **[4M]**

**B).** Mention and discuss the applicability of regression models considering the following domains. In such cases mention the features and response variable. **[4M]**

a) e – commerce b) Pharma industry c) Manufacturing industry d) OTT Platforms

**Q.2**.Consider the following data. **[12M]**

(x, y) = (0,1),(1,3),(2,4),(3,5),(4,7),(6,10),(5,8)

1. Find a suitable cost function which minimizes the cost by considering regression algorithm and justify the selection.
2. If two points (70,150) and (150,250) are added to the data, then what steps you consider to have regression model as a data scientist.
3. Do you suggest any other better model with more applicability? Discuss in detail.

**Q.3.A).**Consider that you are trying to fit a linear regression model to be used for prediction on a given data. 75% of the data is used to train the algorithm and the rest for testing. Write your observations and the plan of action by considering the following situations. **[8M]**

a) Training error is 0.9 and testing error is 0.3

b) Training error is 0.5 and testing error is 0.8

c) Training error is 0.85 and testing error is 0.80.

**B).** Consider a uni-variate regression problem with ‘N’ training examples. The regression model obtained after minimizing sum of squares of error is y = 6.25 + 0.021 x. As the w1 is estimated to be 0.021, which seems relatively smaller value, can we conclude upfront that y is actually not dependent on x. Support your answer with an appropriate reasoning? **[7M]**

**Q.4.A).** How to handle under fitting with respect to regression algorithm? **[2M]**

**B)**.Suppose that a linear regression algorithm is over fitted. Can we proceed with nonlinear regression model to handle it in such case? Discuss it in details. **[3M]**

. **XXXXXXXX**