VII SEMESTER B.TECH. (CHEMICAL ENGINEERING)

END SEMESTER EXAMINATIONS, DEC-JAN 2021

SUBJECT: Machine Learning in Chemical Engineering [CHE4073]

REVISED CREDIT SYSTEM

Date:20 /12/2021 Time duration : 75 minutes (9:20 AM to 10:35 AM) MAX. MARKS: 20

Instructions to Candidates:

- Answer **ALL** questions.
- Missing data may be suitably assumed.

<table>
<tr><td>1A</td><td>Ten data points were taken in an experiment in which the independent variable x is the mole percentage of a reactant and the dependent variable y is the yield (in percent):

<table>
<tr><td>x</td><td>20</td><td>20</td><td>30</td><td>40</td><td>40</td><td>50</td><td>50</td><td>60</td><td>70</td></tr>
<tr><td>y</td><td>73</td><td>78</td><td>85</td><td>90</td><td>91</td><td>87</td><td>86</td><td>75</td><td>65</td></tr>
</table>

Fit a quadratic model of the form $h_\theta(x) = \theta_1 x + \theta_2 x^2$ with these data and estimate the parameter vector using a closed-form linear regression solution. Determine the value of x that maximizes the yield.</td><td>4</td></tr>
<tr><td>1B</td><td>It is decided to design a soft-sensor for a multicomponent distillation column to estimate the product composition and synthesize feedback control design. Explain the detailed step-by-step procedure to design the soft sensor using a suitable machine learning algorithm.</td><td>3</td></tr>
<tr><td>1C</td><td>Why cannot we use the mean square error cost function used in linear regression for logistic regression?</td><td>3</td></tr>
<tr><td>2A</td><td>What is the role of C in the Support Vector Machine? What is the intuition of a large margin classifier?</td><td>3</td></tr>
<tr><td>2B</td><td>How do you ensure you’re not overfitting with a model? What’s the trade-off between bias and variance in the machine learning algorithm</td><td>3</td></tr>
<tr><td>2C</td><td>Reduce the dimension for the given data set below using principal component analysis. Find the principal component (z_1), which captures the maximum variation in the data for the given data set, and determine the percentage variance captured by each principal component. Determine the approximate value of original data using only one principal component (i.e., z_1) and report the original and approximate data error.

<table>
<tr><td>x</td><td>2.5</td><td>2.3</td><td>2.2</td><td>2</td><td>1</td><td>1.5</td><td>1.1</td></tr>
<tr><td>y</td><td>2.4</td><td>2.7</td><td>2.9</td><td>1.6</td><td>1.1</td><td>1.6</td><td>0.9</td></tr>
</table>
</td><td>4</td></tr>
</table>