TECHNO MAIN SALT LAKE

(FORMERLY TECHNO INDIA, SALT LAKE)

Name ..... Bishal Das .....

Roll No. ..... 13030822015 ..... Stream ..... CSE (AIML) .....

Subject ..... Application of ML in Industries (PCC-AIML601) ..... Semester ..... 6th .....

Invigilator's Signature ..... Date .....

PART - A

1. The two most common supervised tasks — (a) Regression, (b) Classification,
2. The purpose of Validation Set is to compare model, on different train-test-splits,
3. In a linear regression problem with a single feature variable, there are two model parameters — (a) Regresson Coefficient & (b) Intercept,
4. The Auc value of a perfect classifier is 1,
5. For a spam email detection system, Precision is more important evaluation metric,

PART - B

6. Train-test-split is the process to split the data into 'Training Set' and 'Test Set', for training and evaluation of the model,

- When the model underperforms on training data, the 'Underfitting' occurs. In the case of 'high-Bias' the model unfits on training data,
- If a model generalizes well on training data but doesn't perform well on new unseen data, then 'Overfitting' occurs,

- How to prevent 'OVERFITTING' :

(i) Reduce noise on data,

(ii) Reduce complexity,

(iii) Use of cross-validation to train-test the model,

(iv) Data cleaning,

(v) Reduce variance on data,

- How to prevent 'UNDERFITTING' :

(i) Adding more data,

(ii) Reduce of Bias-ness,

(iii) Data cleaning,

---

7. The 'Generalization error' is the factor in statistics and Machine Learning to improve the model performance, this generalization error is included ~~three~~ key terms :

(a) BIAS : This is termed as 'how much the model has biasness with noise in data', A high-bias model underfits on training dataset,

(b) VARIANCE : Variance termed as, 'how much the model changes it's predictions with the variations in the data',

A high variance model overfits on training data,

- How to reduce Bias and Variance :

(i) Reducing model complexity results in reducing the variance but it generally increase the biasness of the model, and vice-versa,

(ii) Although, data cleaning, reducing noise in the data & reduces the bias-variance,

- Bias-variance Tradeoff :

When we reduce the complexity of the model, the bias is increased and variance decreases,

$$\text{Variance} \propto \frac{1}{\text{Bias}}$$

This is called Bias-Variance Trade-off.

---

8. The cost functions associated with linear regression,

(i) MAE, (ii) ~~MAE~~ RMSE

- The cost function associated with ~~classification~~ logistic Regression,

CROSS-Entropy,

- General Algorithms that are available to minimize the cost-functions,

⇒ Regularization

→ Ridge Regularization (L2 Regularization)

→ LASSO Regularization (L1 Regularization)

---

9. CONFUSION MATRIX : Confusion matrix is the visualization metric or tool to visualize the model evaluation of a classifier,

- Confusion metrix plays important role in evaluation and see the model predictions, as if is concluded with TP, FP, TN, FN

Predicted values \ Actual values		
	TP	FP
	TN	FN

- $\text{Precision} = \frac{10}{10+82}$

$= 10/92$