**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**
**Indian Institute of Technology Jodhpur**
**Paper Title: Machine Learning 1 ( Minor 1)**

**Time: 1 hour** **Max Marks: 40+4 Bonus**

---

Q1) Suppose we have three categories with P (ω1) = 1/2, P (ω 2 ) = P (ω 3 ) = 1/4 and the following distributions **Marks: 15**

p(x|ω 1 ) ~ N (0, 1)

p(x|ω 2 ) ~ N (.5, 1)

p(x|ω 3 ) ~ N (1, 1) ,

and that we sample the following four points: x = 0.6, 0.1, 0.9, 1.1. Calculate explicitly the probability that the sequence actually came from ω1, ω3, ω3, ω2. Be careful to consider normalization.

Q2) Given the training data, build a decision tree by using Information Gain as the criteria for splitting.
**Marks: 10**

| **Past Trend** | **Open Trend** | **Trading Volume** | **Return** |
|---|---|---|---|
| Positive | Low | Low | Down |
| Negative | High | Low | Down |
| Positive | Low | High | Up |
| Positive | High | High | Up |
| Negative | Low | Low | Down |

Also predict the return of the following new example

Past Trend = Positive , Open Interest = High, Trading Volume = Low

Q3) Highlight the Similarities and Differences between Bootstrap Aggregation and Boosting by giving suitable examples. **Marks: 5**

Q4) Answer the following questions? **Marks: 4**

1. What is the difference between Euclidean distance and Mahalanobis distance?
2. When are the two distances equal?

3. When is it more appropriate to use the Mahalanobis distance over the Euclidean distance?

Q5) Justify your answers with valid reasons for the following questions:- **Marks: 6**

1. What cross validation techniques are suitable for time series analysis ?
2. Your classification model for the cancer detection dataset gives 96% accuracy. Would you like to deploy your model in real life scenarios ?
3. People who bought this also bought…' recommendations seen on Amazon is based on which algorithm?

Q6) Write True or False with justification **Bonus: 4**

1. Cross-validation can help identify overfitting and underfitting.
2. With k-fold cross validation, smaller k is always better.
3. In Bayes classification, the number of actions has to be greater than the possible states of nature.
4. Bagging reduces the bias and boosting reduces the variance.