Topics for CIA-1

Prescribed Textbook

1. **Kevin P. Murphy,** “Machine Learning: A Probabilistic Perspective”, The MIT Press ,2012
2. **C. Bishop**, Pattern Recognition and Machine Learning, Springer, 2007.
3. **Jiawei Han**, Micheline Kamber, and Jian Pei. Data Mining: Concepts and Techniques, Morgan Kaufman, Third Edition, 2012

We are in 2024!

For easier illustration and examples, you may ask ChatGPT, GeekforGeeks or any other source. I am not expecting anything that is not discussed in the class. All the best!

Disclaimer!!

If I ask questions outside of the following list, you must not hold me accountable !!

---

1. **Basics of Machine Learning (Chapter 1 of Kevin P. Murphy, Lecture slides)**
    a. What are the types of Machine Learning models?
    b. What is the no-free lunch theorem?
    c. How do you select your model? Can you provide some examples?
    d. How to split the dataset?
    e. What is meant by training/testing/validation set?
    f. What is meant by loss, bias, variance, error, model complexity?
    g. What is overfitting, under-fitting?
    h. What is meant by hyper-parameter tuning?
    i. What is model deployment?
    j. The curse of dimensionality
    k. Principal component analysis

2. **All about data (Chapters 1-3 of Jiawei Han)**
    a. Data sources, attributes, stats, visualization methods?
    b. Name some data reduction, cleaning, integration, transformation methods.

3. **Supervised/Unsupervised learning.**
    Explain the working principles of (problem statement, math behind the models, working of the model, give example) **(Chapter 1 of Kevin P. Murphy) (Sections 3.1, 3.2 of C.Bishop, Section 9.5.1 of Jiawei Han)**
    a. K-means algorithm
    b. K-nearest neighbor algorithm
    c. Linear Regression
    d. Logistic Regression

4. **Optimization, Evaluation methods (check out the slides, ipython notebook, section 5.2.4 C.Bishop)**
    a. What are some of the ML evaluation metrics?
    b. What are some of the optimization techniques?
    c. Explain how Gradient Descent Algorithm is used in Linear Regression.

5. **Frequent Itemset, Closed, Maximal (Reference: Chapter 6 of Jiawei Han)**
    a. What are frequent itemset?
    b. How is a frequent itemset different from closed and maximal itemset? Provide examples.
    c. What is meant by support, confidence, min.support threshold?
    d. Why aren't certain associations interesting? How to overcome those?
    e. What are other pattern evaluation metrics? How are they useful?
    f. What are the three methods to find the associations in frequent itemset?
    g. Explain the working principles of
        i. Apriori algorithm
        ii. ECLAT algorithm
        iii. FP-Growth
    h. Explain how these algorithms differ for closed and maximal itemset?

6. **Classification methods**
    Explain the working principle of **(Reference: Chapter 8 of Jiawei Han)**
    a. Decision Tree
    b. Random Forest
    c. Basic concepts in classification

7. **Artificial Neural Network** - Questions from the paper you studied for Reading Assignment, Some basic concepts will be covered in the class.
    (**Reference: Chapter 5 of C. Bishop)**

8. If time permits, we will include **Support Vector Machines.**