450-4049/04

Applied Artificial Intelligence Methods

(2024/2025 LS)

Dear student,

Please answer the questions as stated in the notebook and present it as a technical report, upload your answers to the Moodle at the end of the class.

Please note that it is a technical report that need to concise and constructive!

In each task indicate what are the results.

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Email address: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**NoteBook A:**

1. **Logistic regression - Diagnosis**

1.1 Have you succeeded to complete the task?

***R=***

* 1. Write the 5 new features values (X and Y) of 10 patients,

and the probability of diabetes of a new patient, also write the value of the features of the new patient.

***R=***

1. **Decision tree – Classifying patients**

2.1 Have you succeeded to complete the task?

***R=***

2.2 What are your patient features?

***R=***

2.3 Which feature is the most important in making the classification decision?

***R=***

2.4 How could you determine this experimentally?

***R=***

1. **Prediction accuracy by Decision trees**

3.1 Have you succeeded to complete the task?

***R=***

3.2 What is the Test Accuracy test\_acc of the model?

***R=***

3.3 Which depth (d) has better performance?

***R=***

3.4 How does the choice of d impact the model's generalization and overfitting?

***R=***

1. **Binary classification with kNN**

4.1 Have you succeeded to complete the task?

***R=***

4.2 What is the Prediction Accuracy Score?

***R=***

4.3 Which k\_val has better performance?

***R=***

4.4 How does the choice of k affect the bias-variance tradeoff in kNN classification?

***R=***

1. **Binary classification with SVM**

5.1 Have you succeeded to complete the task?

***R=***

* 1. At what values of mb, the point "X" is assigned to the same class when you used different separating line?

***R=***

5.3 What are the best parameters for d?

***R=***

5.4 What happen to the boundaries when you run the “non-linear” separable data plot?

***R=***

5.5 What happen to the data plot when you try different angles?

***R=***

5.5 If you want to separate the two classes in a higher dimension linearly, which angle will you use to draw a better line?

***R=***

5.6 Is it possible to use linear or even kernel based SVM to classify these data?

***R=***

5.7 What are the difference in the two plotted figures?

***R=***

1. **Random Forrest**

6.1 Have you succeeded to complete the task?

***R=***

6.2 Did the Random Forest model outperformed the Decision Tree in terms of classification performance in the same data.?

***R=***

6.3 How does the complexity of the classification problem, can be raised ?

***R=***

1. ***K-means Clustering***

***7.1 Are you able to understand the how k-means clustering?***

***R=***

* 1. ***What will the effect of changing the value of k, in the k-means clustering?***

***R=***

* 1. ***Does increasing the number of principle components has an impact on the performance of clustering?***

***R=***

1. ***Hierarichal Clustering***
   1. ***Write a code to implement divisive clustering?***
   2. ***Comment on the performances of agglomerative and divisive clustering. Which one is best and why?***