

Practice 8- Supervised learning: Movie Profit Prediction Model Comparison

Background: You want to predict whether movies will be profitable based on their screenplays. You hire two critics, A and B, to read a script and rate it on a scale of 1 to 5. The critics are not perfect; here are five data points including the critics' scores and the performance of the movie:

Movie Name	A Score	B Score	Profit?
Pellet Power	1	1	No
Ghosts!	3	2	Yes
Pac is Bac	4	5	No
Not a Pizza	3	4	Yes
Endless Maze	2	3	Yes

Tasks:

1. Data Visualization and Linear Separability Analysis:

- Plot the data points on a 2D plane, label profitable movies with '+' and non-profitable movies with '-', and determine if the data are linearly separable.
- Analyze which model, SVM, NN, or DT, might be more suitable in this situation and explain why.

2. Perceptron Model Training and Feature Selection:

- Train a perceptron model using the given scores as features, including a bias feature ($f_0=1$, $f_1=A$'s score, $f_2=B$'s score).
- Describe how to update the weights and provide the weights after the first update.
- Discuss whether the perceptron model can perfectly classify the following scenarios with the given features and explain why:
 - (i) If the total of their scores is more than 8, the movie will definitely be a success; otherwise, it will fail.
 - (ii) The movie will succeed if and only if each critic gives either a score of 2 or a score of 3.
 - (iii) The movie will succeed if and only if both critics agree.

3. Soft Margin SVM Analysis:

- Given that the data is not linearly separable, apply a soft margin SVM to classify the data.
- Explain the role of the regularization parameter C in the context of soft margin SVM and how it affects the decision boundary.
- Discuss the trade-off between maximizing the margin and minimizing the classification error in the context of soft margin SVM.

4. Decision Tree Model Construction:

- Build a decision tree model using the same dataset and discuss how to choose the appropriate splitting features.

- Analyze the impact of the depth of the decision tree on model performance and discuss how pruning can prevent overfitting.

5. **Model Comparison and Selection:**

- Compare the advantages and disadvantages of SVM, NN, and DT in handling this dataset.
- Provide a comprehensive evaluation, explaining which model to choose in practical applications and the reasons for choosing it.