

Introduction to Artificial Intelligence

Teamwork – Machine Learning

Dec 6, 2023

Objectives

Practice and get familiar with learning from observations. In this assignment you need to make use of the taught subject matters about Machine Learning (ch. 18 and 18.7).

Program

Write programs to build a prediction model using the training data:

- Use **decision tree** or **neural network** (or others)
- Training data (`train.csv`):
 - #examples: 500,000 (218 MB)
 - #attributes (input): 35
 - #classes (output: 0/1): 2
 - imbalanced: #0s > #1s
- Write a program to generate a hypothesis h and use it to predict test data. To “test” h , your program must
 - input test data from `test.csv` file
 - **Kaggle submission**¹
 - * Sample submission
 - * Final test
- Performance indexes:
 - Accuracy = $\frac{tp+tn}{tp+tn+fp+fn}$
 - Precision = $\frac{tp}{tp+fp}$
 - Recall = $\frac{tp}{tp+fn}$
 - **F-score** = $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

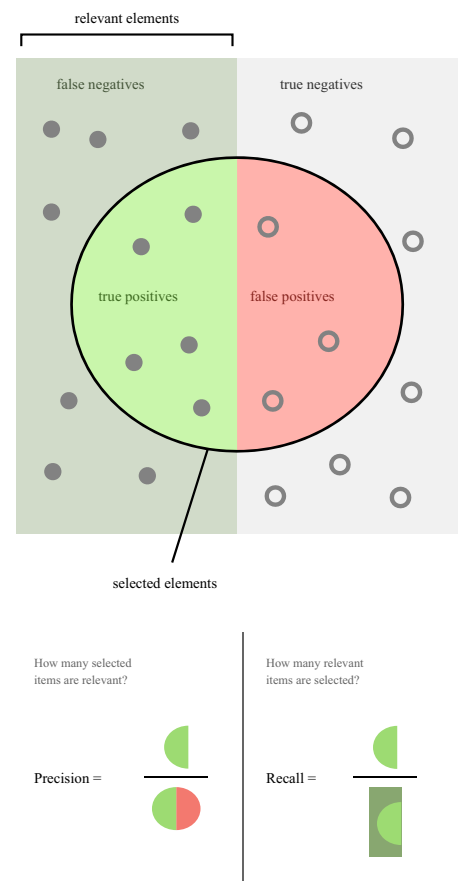


Figure 1: Precision and Recall

¹<https://www.kaggle.com/t/881442ed177c4ada81efae38e0b61b7b>

Report

- Explain your methods
- Describe how you improve the results
- Evaluate each of your teammates
(ps. each team consisting of 1, 2, or 3 people)
 - Score: -5 (bad) – 5 (good)
 - Give short comment for *each* teammate

Evaluation

- Report: 60%
- Performance: 60% (final test's Accuracy and F-score: 30%; rank: 30%)
- Cooperation (teammates' evaluation): 30%

Submission

- **2023/12/22 24:00 (degrade by 10 points for each day delay)**
- Source code + Report (no longer than **six** A4 pages)
- Zip the source code and report. Upload it to eeclass.

Policy on AI-Generated Contents

The use of AI-generated contents (e.g., by ChatGPT) in any work done on assignments and projects shall be disclosed. Each part of the work that uses AI-generated contents shall have a citation to the AI system used to generate the contents.

Using any AI tool or system at exams is forbidden.