

# Introduction to Artificial Intelligence

## Program – Machine Learning

Apr 28, 2025

### 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. 19 and 22.1).

### 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: 156,076
  - #attributes (input): 96
  - #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**<sup>1</sup>
    - \* 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** =  $2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

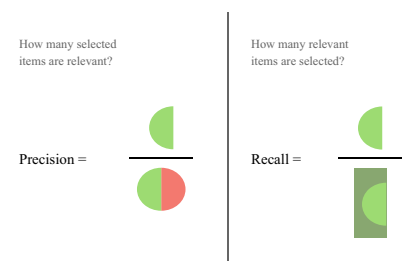
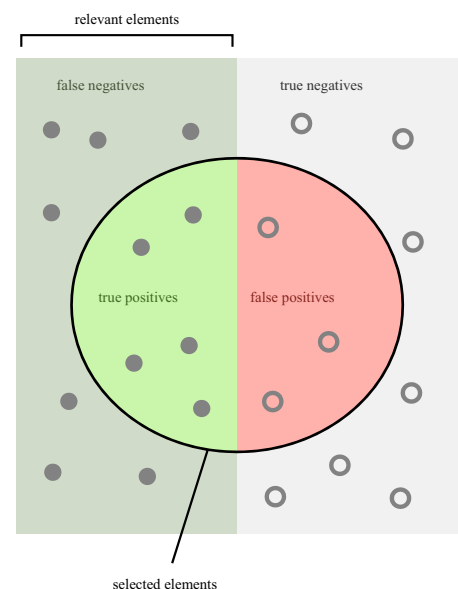


Figure 1: Precision and Recall

### Report

- Explain your methods
- Describe how you improve the results

<sup>1</sup><https://www.kaggle.com/t/e6a7d20bd9114278a3574c4359383c0a>

## Grading Policy

- Report and correctness: 70%
  - Model's performance must exceed baseline: 30%
  - Report quality: 40%
- Performance: 30%
  - Rank (public): 15%
  - Rank (private): 15%

## Submission

- **2025/05/14 24:00 (degrade by 10 points for each day delay)**
- Source code + Report (no longer than **six** A4 pages)
- Upload the source code and report to eclass.

## Policy on AI-Generated Contents

*Work created by AI tools may **not** be considered original work and instead, considered **automated plagiarism**. It is derived from previously created texts from other sources that the models were trained on, yet doesn't cite sources.*

*A responsible use of AI-based tools in completing coursework must be done in accordance with the following:*

- 1. You must clearly identify the use of AI-based tools in your work. Any work that utilizes AI-based tools must be clearly marked as such, including the specific tool(s) used. For example, if you use ChatGPT-4o, you must cite "ChatGPT-4o. (YYYY, Month DD of query). "Text of your query." Generated using OpenAI. <https://chat.openai.com/>"*
- 2. You must be transparent in how you used the AI-based tool, including what work is your original contribution. An AI detector may be used to detect AI-driven work.*
- 3. You must not use AI-based tools to generate the content of your work, but you may use them for editing.*
- 4. You must ensure your use of AI-based tools does not violate any copyright or intellectual property laws.*
- 5. You must not use AI-based tools to plagiarize without citation.*