

# Introduction to Artificial Intelligence

## Project 2 – Classification & Clustering

Jianmin Li

Department of Computer Science and Technology  
Tsinghua University

Spring, 2024

# Supervised and Unsupervised Learning

- Training with MNIST

0	0	0	0	0
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9

# Pre-requirements

- At least 400MB disk space and 800MB memory
- Python package that you should be familiar with
  - numpy
  - sklearn
- Python packages that should be installed:
  - numpy
  - skimage
  - sklearn
- Anaconda is recommended

# Basic Tasks (1)

- K-Means (4 points)
  - Implement **KMeansCluster.fit** in **featureExtractor.py**
  - `python featureExtractor -f kmeans -s 10`
- KNN (3 points)
  - Implement **KNNClassifier.classify** method in **classifiers.py**
  - `python dataClassifier.py -c knn -n 5`
- Softmax Regression (4 points)
  - Implement **PerceptronClassifier.train** in **classifiers.py**
  - `python dataClassifier.py -c perceptron`
- sklearn **MUST NOT** BE USED in the **above tasks**, OR you **will not** pass the autograder

# Basic Tasks (2) & Bonus

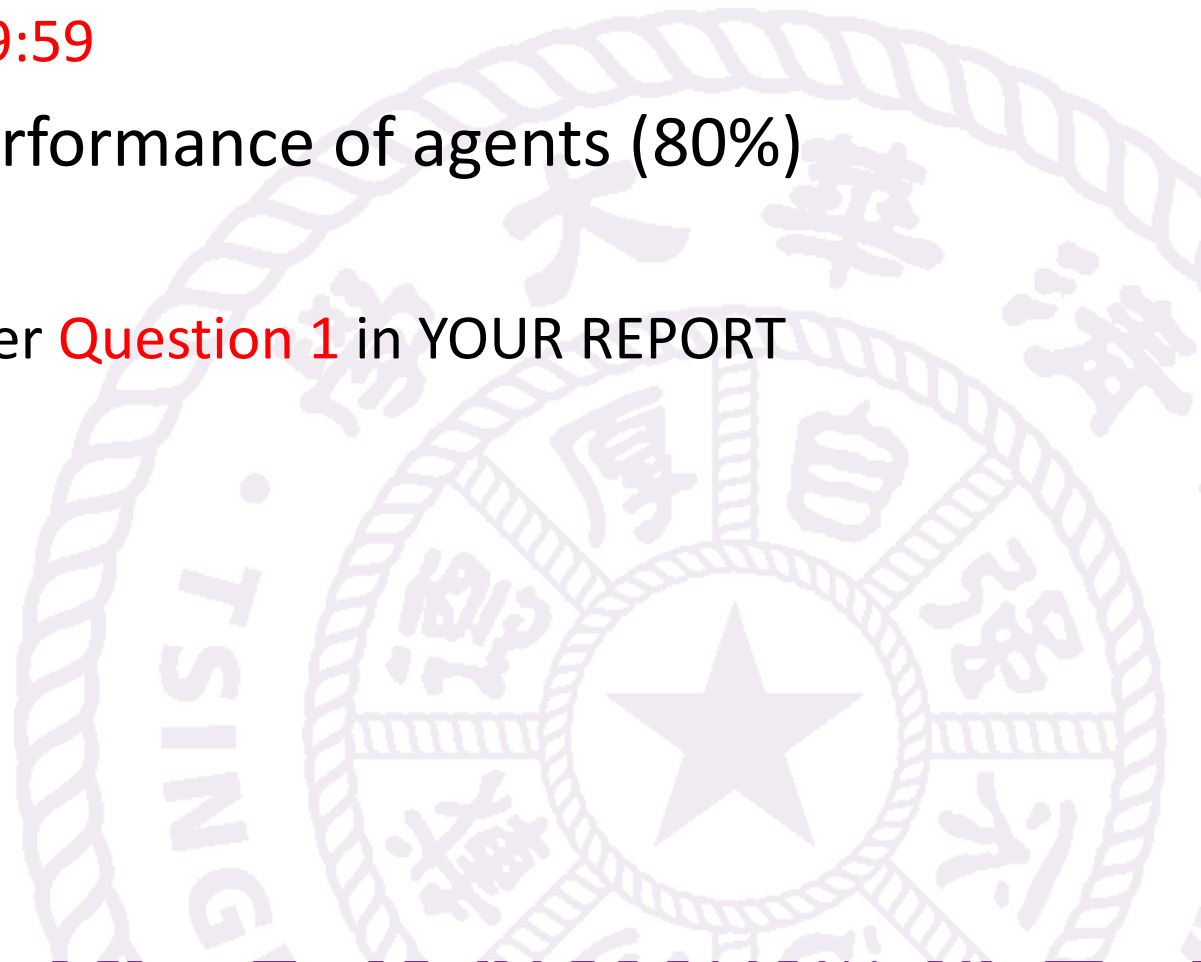
- Training SVM with sklearn (2 points)
  - Implement `SVMClassifier.train`, `SVMClassifier.classify` using package sklearn, in `classifiers.py`
  - You should be familiar with some sklearn API
- Obtaining better classification results (2 points + 1 point bonus)
  - Implement `BetterClassifier.train`, `BetterClassifier.classify` in `classifiers.py`
  - You may make use of sklearn package
  - Try to obtain good accuracy as much as you can
  - **1 point extra credit** for the leading classification accuracy

# Submission

- A 1-3 pages report (either Chinese or English)
  - You **MUST** answer **Question 1** in YOUR REPORT
  - You will not get full report credits if cannot answer the above questions correctly
  - Some analysis on different algorithms/feature extractor techniques is useful for better grading
- Zip the files as the following structure
  - student\_id.zip (e.g. 20090112xx.zip)
  - student\_id.pdf
  - classifiers.py
  - featureExtractor.py

# Grading

- Due
  - 2024/5/29 23:59:59
- Correctness & performance of agents (80%)
- Report (20%)
  - You **MUST** answer **Question 1** in YOUR REPORT



谢谢！

