**L01-intro summary**

# What is Pattern Recognition?

Pattern recognition เป็นสาขาของ Machine Learning ที่ศึกษาเกี่ยวกับการหา ความสม่ำเสมอของลักษณะรูปแบบภายในข้อมูล

# Machine Learning (ML), Pattern Recognition (PR), Data Mining (DM) และ Knowledge Discovery in Databases (KDD) ต่างกันอย่างไร?

ไม่ต่างเพราะ แต่ละศาสตร์ล้วนสนใจว่า เราได้เรียนรู้อะไรจากข้อมูล ต่างกันที่มุมมองว่า เราจะออกแบบเครื่องมือ หรือเราจะนำเครื่องมือไปใช้

สร้างเครื่องมือ

ใช้เครื่องมือ

DM PR ML AI

แต่ละศาสตร์จะมีหัวข้อที่แตกต่างกันบางเรื่อง

DM – Data warehouse, ETL

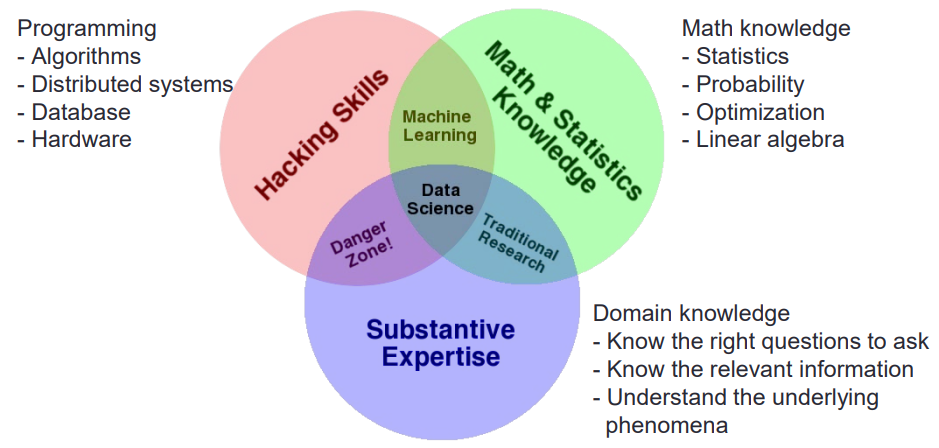
AI – search, swarm intelligence

PR – Signal processing (feature engineering)

# Terminologies

|  |  |
| --- | --- |
| Machine learning | Statistics |
| Network, Graphs | Model |
| Weights | Parameters |
| Learning | Fitting |
| Generalization | Test set performance |
| Supervised learning | Regression/Classification |
| Unsupervised learning | Density estimation, Clustering |
| Large grant = $1,000,000 | Large grant = $50 ,000 |

# Data Science & Data Analytics



# Types of machine learning

1. Supervised learning: Learn a model F from pairs of (x, y)
2. Unsupervised learning: Discover the hidden structure in unlabeled data x (no y)
3. Reinforcement learning: Train an agent to take appropriate actions in an environment by maximizing rewards

# Typical workflow of machine learning

1. Feature extraction (getting the x)
2. Modeling

* Training (getting the function F)

1. Evaluation

* Metrics (defining what’s the best function F) สิ่งที่ใช้พิจารณา model ว่าดีไม่ดี (Errors/failures, Accuracy/success) เทียบกับคำตอบบน Ground truths
* Testing (getting the y for unseen inputs)

# Evaluation

## Confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
|  | | **Predicted Class (Detector)**  **(ค่า y’ ที่ model ทำนายได้)** | |
| Yes | No |
| **True Class (Actual)**  **(ค่า y ของ data x)** | Yes | True positive | False negative  (Type II error) |
| No | False positive  False Alarm  (Type I error) | True negative |

## Definitions

* True positive rate (Recall, sensitivity) = # true positive / # of actual yes
* False positive rate (False alarm rate) = # false positive / # of actual no
* False negative rate (Miss rate) = # false negative / # of actual yes
* True negative rate (Specificity) = # true negative / # of actual no
* Precision = # true positive / # of predicted positive
* F score (F1 score, f-measure) เป็น Harmonic mean ของ precision and recall