

## Conclusion

- AI Overview
    - History AI
    - AI & Machine Learning
    - Machine Learning
      - Supervised Learning ⇒ การวิเคราะห์/การพยากรณ์
        - Regression
        - Classification
      - Unsupervised Learning ⇒ การทำความเข้าใจโครงสร้างของข้อมูล
        - Clustering
        - Dimensionality Reduction
        - Association
        - Generative Model (ChatGPT & MidJourney)
      - Reinforcement Learning ⇒ การเรียนรู้แบบลองผิดลองถูก
        - Policy
    - Supervised Learning
      - Data ⇒ Model ⇒ Prediction
    - **Supervised Learning Workflow**
      - Data ⇒ ready-to-use (numerical & table)
      - Model Creation
      - Model Evaluation
-

- Data Preparation
    - Data Stating
      - การแนะนำ data แบบ formal
    - Data Requirement
      - Numerical Data
      - Table
    - Table (Structured Data)
      - Data Type
        - Numerical Data
        - Categorical Data
          - Ordinal Data (มีลำดับ)
          - Nominal Data (ไม่มีลำดับ)
      - Ordinal Encoding
      - One Hot Encoding
      - Q1 : ข้อมูลที่เป็น Categorical Data มันสามารถตัวเลขเป็นได้ไหม?
      - A : เป็นไปได้
        - ถ้าเป็นแบบ Ordinal Data มันจะไม่ใช่ไร ⇒ สามารถใช้งานได้เลย
        - ถ้าเป็นแบบ Nominal Data ⇒ One Hot Encoding
      - Q2 : เราจะรู้ได้อย่างไรว่าข้อมูลเป็นแบบ Ordinal Data หรือ Nominal Data?
      - A2 : ความเข้าใจในข้อมูล ซึ่งแต่ละคนอาจมองไม่เหมือนกันก็ได้
      - \*\*\* วิธีการ Encoding จะทำให้ดีต้องเข้าใจ algorithm ที่ใช้ในการสร้าง model ด้วย
    - Not Table (Unstructured Data)
      - Image
        - Read Image ⇒
        - Resize Image ⇒
        - Flatten ⇒
        - Vstack ⇒ Table
      - Text
        - Count Vectorization ⇒ Table
      - Sound
        - Read Sound ⇒
        - Transform to Image ⇒
        - Resize Image ⇒
        - Flatten ⇒
        - Vstack ⇒ Table
-

- Model Evaluation
    - Regression
      - R-squared
        - $< 0 \Rightarrow$  แย่กว่าพยากรณ์ด้วยค่าเฉลี่ยของ  $y$
        - $= 0 \Rightarrow$  พยากรณ์ได้ดีเท่าค่าเฉลี่ยของ  $y$
        - $> 0 \Rightarrow$  ดีกว่าพยากรณ์ด้วยค่าเฉลี่ยของ  $y$
        - $= 1 \Rightarrow$  perfect
      - Mean Squared Error
      - Mean Absolute Error
      - Mean Absolute Percentage Error
    - Classification
      - Confusion Matrix
      - Accuracy  $\Rightarrow$  จำนวน sample ที่พยากรณ์ถูก / จำนวน sample ทั้งหมด
        - การใช้งานจะไม่ specific
        - ง่าย
        - ใช้งานได้ไม่ดีกับ imbalanced class
      - Precision  $\Rightarrow TP / TP + FP \Rightarrow$   
จำนวน positive ที่พยากรณ์ถูกต้อง / จำนวน positive ทั้งหมดที่พยากรณ์ออกมา
      - Recall  $\Rightarrow TP / TP + FN \Rightarrow$   
จำนวน positive ที่พยากรณ์ถูกต้อง / จำนวน actual ที่เป็น positive ทั้งหมด
      - Precision VS Recall
        - ผู้พิพากษา
        - หมอตรวจมะเร็ง
      - F1 score  $\Rightarrow$  ค่าเฉลี่ยแบบ harmonic mean ของ precision & recall
        - การใช้งานจะไม่ specific
        - ยาก
        - ใช้งานได้ดีกับข้อมูลที่เป็น imbalanced class
-

- 12 Algorithm
  - Linear Regression
    - คืออะไร : สมการเชิงเส้นที่ใช้ตัวแปรต้นเพื่อพยากรณ์ตัวแปรตาม
 
$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p$$
    - แก้ปัญหา Regression
    - การสร้าง model  $\Rightarrow$  minimize SSE โดยการใช้ least squared method
    - ข้อดี
      - ความเป็นาง่าย
      - ไม่ค่อย overfit
    - ข้อเสีย
      - ใช้ได้ดกกับแค่ข้อมูลที่มีพฤติกรรมแบบเชิงเส้น
      - การเข้าใจ algorithm นี้ให้ลึกเป็นเรื่องยากมาก (โดยเฉพาะในส่วนของ assumption)
    - การด้อยอด
      - Ridge Regression
      - Lasso Regression
      - Elastic Net
  - Logistic Regression
    - คืออะไร : การด้อยอดสมการเชิงเส้นเพื่อบ่ง class ของข้อมูล
 
$$\hat{y} = \sigma(w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p)$$
    - \*\*\* 2-class vs multi-class
    - แก้ปัญหา Classification
    - การสร้าง model  $\Rightarrow$  minimize Cross Entropy โดยการใช้ gradient descent
    - ข้อดี
      - ความเป็นาง่าย
      - ไม่ค่อย overfit
    - ข้อเสีย
      - ใช้ได้ดกกับแค่ข้อมูลที่มีพฤติกรรมแบบเชิงเส้น
      - การเข้าใจ algorithm นี้ให้ลึกเป็นเรื่องยากมาก (โดยเฉพาะในส่วนของ assumption)
    - การด้อยอด
      - L2 Regularization / L2 penalty
      - L1 Regularization / L1 penalty
      - Elastic Net
      - Dealing Imbalanced Class
  - Neural Network (1 hidden layer)
    - คืออะไร : linear function  $\Rightarrow$  nonlinear function  $\Rightarrow$  หลาย ๆ nf มาประกอบร่างกัน  $\Rightarrow$  เครื่องมือสำหรับ approximate nonlinear function ที่มีความซับซ้อน
    - แก้ปัญหา Regression & Classification
    - การสร้าง model
      - Regression  $\Rightarrow$  minimize MSE โดยการใช้ gradient descent

- Classification  $\Rightarrow$  minimize Cross Entropy โดยใช้ gradient descent
- ข้อดี
  - เป็น algorithm ที่ทรงพลังมาก (approximate ได้ทุก function ในโลกนี้)
- ข้อเสีย
  - Overfit
  - High computational cost
- Deep Learning (>1 hidden layer)
  - คืออะไร : Neural Network ที่มีการเพิ่ม hidden layer เข้ามา  $\Rightarrow$  การ reuse สิ่งที่ hidden layer ในชั้นก่อนหน้าสร้าง  $\Rightarrow$  ในความซับซ้อนที่เท่ากัน DL จะลด computational cost ลง
  - แก้ปัญหา Regression & Classification
  - การสร้าง model
    - Regression  $\Rightarrow$  minimize MSE โดยใช้ gradient descent
    - Classification  $\Rightarrow$  minimize Cross Entropy โดยใช้ gradient descent
  - ข้อดี
    - เป็น algorithm ที่ทรงพลังมาก (approximate ได้ทุก function ในโลกนี้)
  - ข้อเสีย
    - Overfit
    - High computational cost
  - ข้อพึงระวัง
    - i) ความหลากหลายของ plane ใน hidden layer ชั้นแรกต้องมากเพียงพอ
    - ii) ความซับซ้อนต้องมากเพียงพอ
    - iii) จะใช้ i กับ ii เท่าที่จำเป็น  $\Rightarrow$  เพื่อป้องกัน overfit
- Classification Tree
  - คืออะไร : algorithm สำหรับแก้ปัญหา classification โดยที่ตัว model จะอยู่ในรูปของกฎเกณฑ์
  - แก้ปัญหา Classification (nature : decision boundary จะขนานกับแกนเสมอ)
  - การสร้าง model  $\Rightarrow$  การตั้งคำถามที่ทำให้ Information Gain มีค่าสูงสุด (ค่อย ๆ ตั้งไปที่ละคำถาม ทีละขั้น ทีละจุด)
  - สมบัติ : มีความเป็น universal approximator เหมือนกับ NN และ DL
  - ข้อดี
    - Model ที่ออกมาเข้าใจได้ง่าย
  - ข้อเสีย
    - มันเสี่ยง overfit
- Regression Tree
  - คืออะไร : algorithm สำหรับแก้ปัญหา regression โดยที่ตัว model จะอยู่ในรูปของกฎเกณฑ์
  - แก้ปัญหา Regression (nature : step function)

- การสร้าง model  $\Rightarrow$  การตั้งคำถามที่ทำให้ SSE มีค่าต่ำที่สุด (ค่อย ๆ ตั้งไปที่ละคำถาม ทีละขั้น ทีละจุด)
- สมบัติ : มีความเป็น universal approximator เหมือนกับ NN และ DL
- ข้อดี
  - Model ที่ออกมาเข้าใจได้ง่าย
- ข้อเสีย
  - มันเสี่ยง overfit
  
- k Nearest Neighbor
  - คืออะไร : algorithm สำหรับแก้ปัญหา classification โดยพิจารณาจากเพื่อนบ้านที่ใกล้ที่สุด k ตัว
  - แก้ปัญหา : Classification & Regression
  - การสร้าง model  $\Rightarrow$  วัดระยะห่าง & เลือก k ตัวมา majority
  - \*\*\* ในปัญหา Regression  $\Rightarrow$  พิจารณาเพื่อนบ้านที่ใกล้ที่สุด k ตัว (หาค่าเฉลี่ย)
  - ข้อดี
    - ง่าย
  - ข้อเสีย
    - Curse of dimensionality
  
- Naive Bayes
  - คืออะไร : algorithm สำหรับแก้ปัญหา classification โดยตั้งต้นคือการพิจารณาความน่าจะเป็นของการเป็น class นั้น ๆ จากค่า feature  $\Rightarrow$  practical  $\Rightarrow$  conditional probability  $\Rightarrow$  naive bayes
  - แก้ปัญหา : Classification
  - การสร้าง model  $\Rightarrow$  คำนวณ prob ตามสูตรของ naive bayes (เย่ ! เรารู้ที่มาสูตรด้วยนะ)
  - ข้อดี
    - ง่าย (ดูได้จากตอนที่เรากำลังทำการคำนวณ)
  - ข้อเสีย
    - Assumption แรงมาก (ไม่สามารถรักษาได้ในชีวิตจริง)
  - ตรรกศาสตร์ : ถ้าตรงตาม assumption  $\Rightarrow$  performance ของ model จะดี
  - ตรรกศาสตร์ : ถ้าไม่ตรงตาม assumption  $\Rightarrow$  จะการันตีไม่ได้ว่า performance ของ model จะดีหรือไม่
  
- Support Vector Classification
  - คืออะไร : สร้าง hyperplane ที่ทำให้ margin มีขนาดกว้างมากที่สุด
  - แก้ปัญหา : Classification
  - การสร้าง model
    - Objective : maximize ขนาดของ vector w ยกกำลัง 2
    - Constraint : ข้อมูลจะต้องแบ่งได้อย่างถูกต้องทุก sample
  - ข้อดี
    - ใช้งานได้กับทั้งข้อมูลที่ซับซ้อน & ไม่ซับซ้อน
  - ข้อเสีย
    - (เครื่องมืออื่น ๆ มา join/plug ด้วยยาก)
    - เป็น algorithm ที่ซับซ้อน และยากต่อการทำความเข้าใจ

- การต่อยอด
  - พื้นฐาน
  - One-to-rest
  - Improve optimization
  - Improve generalization
  - Improve performance + extend scope
  - \*\*\* Lagrange Multipliers + KKT
- Support Vector Regression
  - คืออะไร : สร้าง hyperplane ที่ทำให้ margin กว้างมากที่สุดภายใต้ constraint ที่กำหนด (บางครั้งอาจหาคำตอบ)
  - การสร้าง model
    - Objective : maximize ขนาดของ vector  $w$  ยกกำลัง 2
    - Constraint : ห้ามความผิดพลาดในการพยากรณ์ของ model เกินเท่าไร
  - ข้อดี
    - ใช้งานได้กับทั้งข้อมูลที่ซับซ้อน & ไม่ซับซ้อน
  - ข้อเสีย
    - (เครื่องมืออื่น ๆ มา join/plug ด้วยยาก)
    - เป็น algorithm ที่ซับซ้อน และยากต่อการทำความเข้าใจ
- Gaussian Process
  - คืออะไร : การพิจารณาความสัมพันธ์ของข้อมูลทุกคู่ใน training set เพื่อที่จะ map กับค่า  $y$
  - แก้ปัญหา : Regression
  - การสร้าง model  $\Rightarrow$  maximum likelihood
  - ข้อดี
    - เป็น algorithm เพียงตัวเดียวที่ให้ค่าความมั่นใจในการพยากรณ์ออกมา (sd)  $\Rightarrow$  ปรากฏใน ensemble method ด้วย
  - ข้อเสีย
    - Overfit  $\Rightarrow$  คนใช้ algorithm นี้ยาก
- Linear Discriminant Analysis
  - คืออะไร : การสร้างมิติใหม่ ที่ทำให้ข้อมูลใน class เดียวกันจะกระจุกตัวกันมากสุด ในขณะที่ข้อมูลคนละ class จะมีระยะห่างกัน (จากจุดศูนย์กลาง) มากที่สุด
  - แก้ปัญหา : Classification
  - ข้อดี
    - ทำได้ทั้ง dimensionality reduction & classification
    - LDA จะไม่ถูกแทนที่ในกรณีที่ เราต้องการ model ตามคำอธิบายข้อมูล
  - ข้อเสีย
    - คณิตศาสตร์ซับซ้อน และยากต่อการทำความเข้าใจ (theory)
    - ถ้าอยากได้แค่สมบัติการเป็น classification เราใช้ logistic regression แทนได้
    - ถ้าอยากได้สมบัติของ dimensionality reduction เราใช้ PCA
    - แต่ถ้าอยากได้สมบัติทั้ง 2 พร้อมกัน

- PCA + algorithm อื่น ๆ ดีกว่า

#### Ensemble Method

1. Boosting
2. Bagging
3. Stacking