

Logistic Regression

SVM

อ. ปรัชญ์ ปิยะวงศ์วิศาล

Pratch Piyawongwisal

Today

- Recap - Linear Regression
 - Minimizing cost function
 - Solutions
 - Normal Equation
 - (Stochastic) Gradient Descent
 - Polynomial Regression
 - Regularization -> solves overfitting
- Logistic Regression

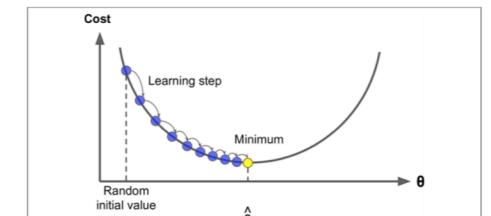
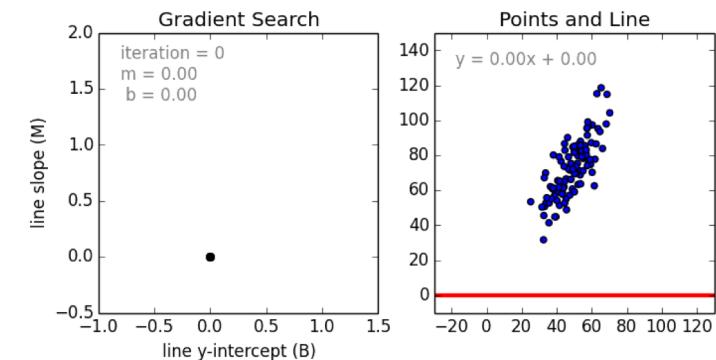
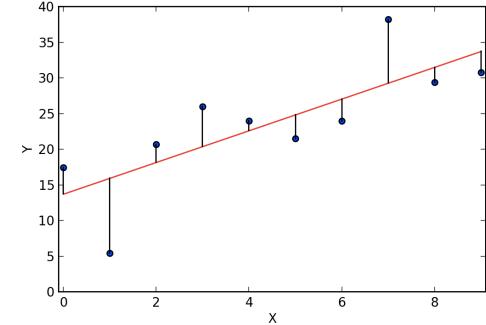
Linear Regression - Summary

- Regression: หาโมเดลที่ fit กับข้อมูลได้ดีที่สุดได้อย่างไร?
- โมเดล: $\hat{y} = h_{\theta}(x) = \theta^T x$
- cost function ของโมเดล: mean-square error (MSE)

$$J(\theta) = MSE(\theta) = \frac{1}{M} \sum_{i=1}^M (\theta^T x^{(i)} - y^{(i)})^2$$

- การ train โมเดล คือ การหาค่าของ $\hat{\theta}_{MSE} = \underset{\theta}{\operatorname{argmin}} J(\theta)$
- Solution 1: แก้สมการตรง ๆ จะได้ Normal equation: $\hat{\theta}_{MSE} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}$
- Solution 2: หรือใช้วิธี Gradient Descend โดยค่อยๆ update θ :

$$\theta^{(\text{next step})} = \theta - \eta \nabla_{\theta} MSE(\theta)$$

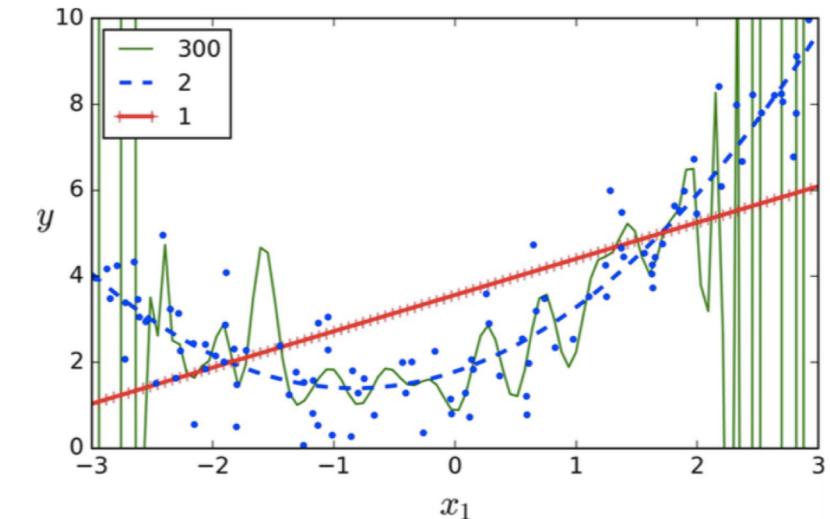


ภูเข้า MSE

Polynomial Regression, Regularization

- สามารถทำ Polynomial Regression โดยเพิ่มมิติข้อมูล training เช่น
 - x_i เดิม = [น้ำหนัก, ส่วนสูง, อายุ] = [70, 150, 30]
 - x_i ใหม่ = [70, 150, 30, 70^2 , 150^2 , 30^2]
- แต่ถ้า poly degree สูงไปอาจทำให้ overfit 😞
- Solution: เพิ่มพจน์ regularization ใน cost function

$$J(\theta) = \text{MSE}(\theta) + \alpha \frac{1}{2} \sum_{i=1}^n \theta_i^2$$



Ridge (L2) regularization

ช่วยลดอิทธิพลของ polynomial degree สูงๆ อย่าง x^4, x^5 = ลด overfitting

Exercise: Housing Price Prediction

- ข้อมูล training

ขนาด (Sq. ft.)	จำนวน ห้องนอน	ระยะทางไป ห้างสรรพสินค้า	ราคา (ล้านบาท)
1000	2	5	9.5
1500	3	30	8.0
2000	5	40	12.5
1700	1	5	9.0
1200	2	30	5.5

กำหนดให้โมเดล Linear Regression คือ

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

โดยที่

x_1 = ขนาดบ้าน (Sq. ft.)

x_2 = จำนวนห้องนอน

x_3 = ระยะทางไปห้างสรรพสินค้าที่ใกล้ที่สุด

- จงหา $\hat{\theta}_{MSE}$ ด้วยวิธี Normal Equation
- จงคำนวณราคาร้านที่มี $X = (1200, 3, 8)$

Exercise: Housing Price Prediction

- ข้อมูล training

ขนาด (Sq. ft.)	จำนวน ห้องนอน	ระยะทางไป ห้างสรรพสินค้า	ราคา (ล้านบาท)
1000	2	5	9.5
1500	3	30	8.0
2000	5	40	12.5
1700	1	5	9.0
1200	2	30	5.5

เราสามารถเขียนข้อมูลในรูป matrix ได้ดังนี้

$$X = \begin{bmatrix} 1 & 1000 & 2 & 5 \\ 1 & 1500 & 3 & 30 \\ 1 & 2000 & 5 & 40 \\ 1 & 1700 & 1 & 5 \\ 1 & 1200 & 30 & 5.5 \end{bmatrix}$$

$$y = \begin{bmatrix} 9.5 \\ 8.0 \\ 12.5 \\ 9.0 \\ 5.5 \end{bmatrix}$$

Recap: Supervised Learning

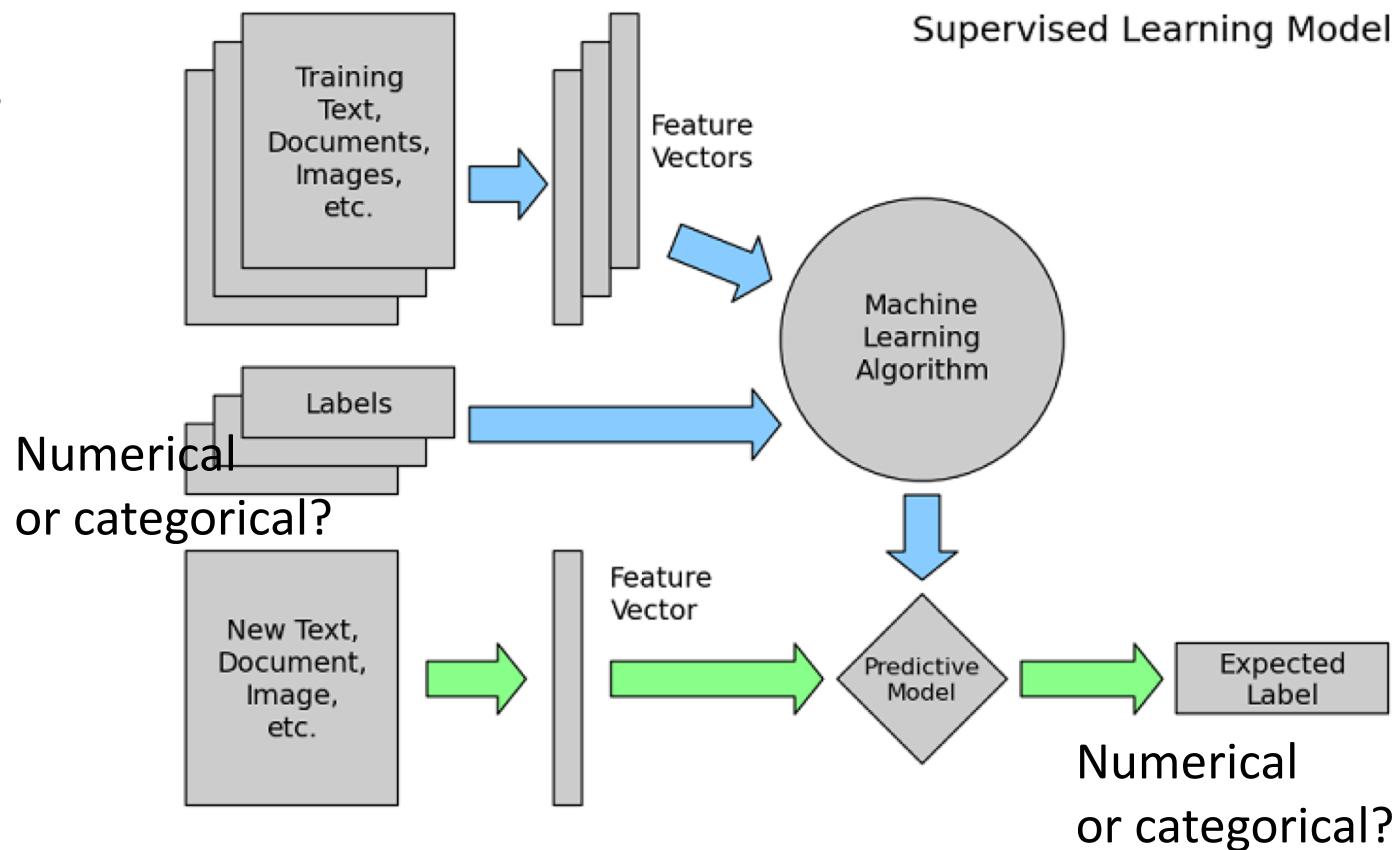
- Classification

kNN

- Predicts class labels/categories
- ทำนายค่าที่เป็นหมวดหมู่ = จำแนกประเภท
- อาจมองเป็นการหา **boundary** ที่แบ่งข้อมูลในแต่ละหมวดหมู่ ออกจากกัน

- Regression

- Linear Regression
- Predicts continuous values
 - ทำนายค่าที่เป็นจำนวนจริง
 - อาจมองเป็นการหา **hyperplane** ที่ fit กับข้อมูลที่มีมากที่สุด



Logistic Regression

- เราสามารถใช้ regression algo ในการทำ **classification** ได้ด้วย !?
- เพื่อความง่าย พิจารณาปัญหา **binary classification** เช่น
 - อีเมลเป็น spam (label=1) หรือไม่เป็นสแปม (label=0)
 - คนไข้เป็นมะเร็ง (label=1) หรือไม่เป็นมะเร็ง (label=0)
- **Key Idea:** เราต้องการให้มีผลลัพธ์ $\hat{y} = h_{\theta}(x)$ ที่นายค่า
 - $p(y = 1|x; \theta) = h_{\theta}(x)$ ความน่าจะเป็นที่ **label** จะเป็น 1
 - $p(y = 0|x; \theta) = 1 - h_{\theta}(x)$ ความน่าจะเป็นที่ **label** จะเป็น 0
- จากนั้นเราสามารถใช้ค่าของ p ในการ **classify** ข้อมูลด้วยกฎง่ายๆ นี้ได้

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5, \\ 1 & \text{if } \hat{p} \geq 0.5. \end{cases}$$

Logistic Regression

- **Key Idea:** เราต้องการให้โมเดล $\hat{y} = h_{\theta}(x)$ ทำนายค่า
 - $p(y = 1|x; \theta) = h_{\theta}(x)$ ความน่าจะเป็นที่ **label** จะเป็น **1**
 - $p(y = 0|x; \theta) = 1 - h_{\theta}(x)$ ความน่าจะเป็นที่ **label** จะเป็น **0**
- Q: ใช้ $h_{\theta}(x)$ เดียวกันกับ Linear Regression เลยได้ไหม?

$$h_{\theta}(x) = \boldsymbol{\theta}^T \mathbf{x}$$

Logistic Regression

- **Key Idea:** เราต้องการให้โมเดล $\hat{y} = h_{\theta}(x)$ ทำนายค่า
 - $p(y = 1|x; \theta) = h_{\theta}(x)$ ความน่าจะเป็นที่ **label** จะเป็น **1**
 - $p(y = 0|x; \theta) = 1 - h_{\theta}(x)$ ความน่าจะเป็นที่ **label** จะเป็น **0**
- Q: ใช้ $h_{\theta}(x)$ เดียวกันกับ Linear Regression เลยได้ไหม?
$$h_{\theta}(x) = \boldsymbol{\theta}^T \mathbf{x}$$
- ปัญหา: ความน่าจะเป็นต้องมีค่าในช่วง 0-1 เท่านั้น ในขณะที่ $\boldsymbol{\theta}^T \mathbf{x}$ มีค่าเป็นเท่าใดก็ได้ 😞

* สังเกตว่าใช้ตัว \hat{p} แทน \hat{y} เพราะเป็นความน่าจะเป็น

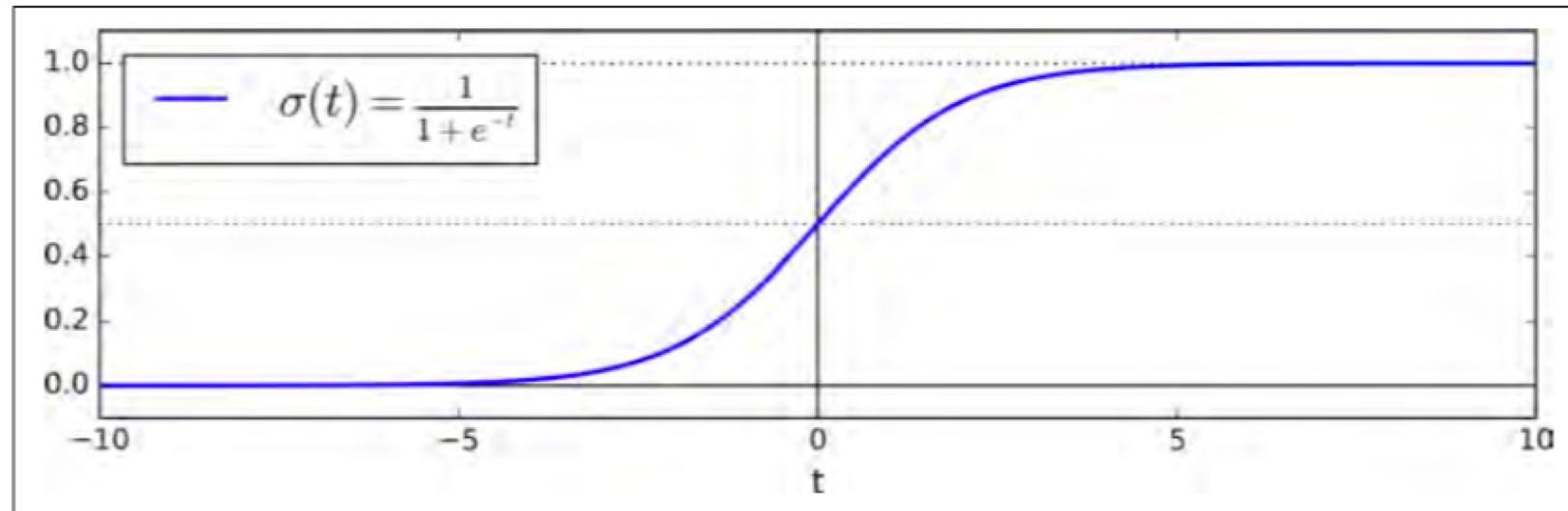
Logistic Regression

- Solution: แค่เปลี่ยน model เป็น

$$\hat{p} = h_{\theta}(x) = \sigma(\theta^T x)$$

- โดยที่ $\sigma(t)$ คือ logistic (sigmoid) function

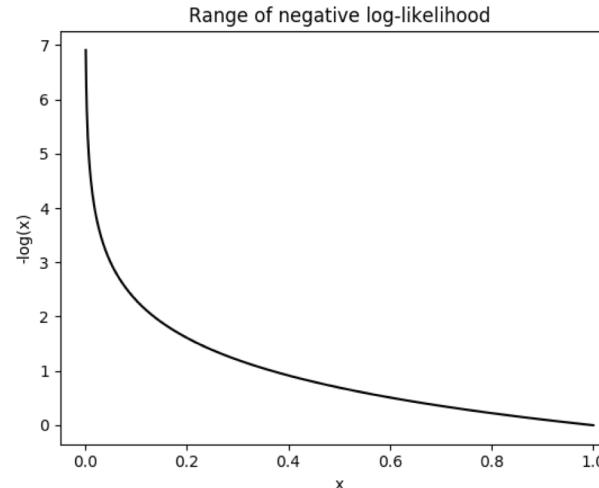
$$\sigma(t) = \frac{1}{1 + \exp(-t)}$$



Logistic Regression

- cost function จะต้องเปลี่ยนไปด้วย
- cost สำหรับแต่ละ instance ข้อมูลคือ

$$c(\theta) = \begin{cases} -\log(\hat{p}) & \text{if } y = 1, \\ -\log(1 - \hat{p}) & \text{if } y = 0. \end{cases}$$



ไอเดียคือ ถ้าเฉลยเป็นมะเร็ง ($y=1$)
แต่ \hat{p} ทำนายว่าใกล้ 0 ค่า cost = $-\log(\sim 0)$ จะใหญ่มาก
ถ้า \hat{p} ทำนายว่าใกล้ 1 ค่า cost จะเป็น 0

- cost สำหรับทั้งชุดข้อมูลจึงเป็นดังสมการ

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)}) \right]$$

Logistic Regression

- จะ train โมเดลนี้อย่างไร? (minimize cost อย่างไร?)

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)}) \right]$$

- ไม่มี closed-form solution แบบ Normal Eq 😞

Logistic Regression

- จะ train โมเดลนี้อย่างไร? (minimize cost อย่างไร?)

[Go to page 158](#) $J(\theta) = \frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(\hat{p}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{p}^{(i)}) \right]$

- ไม่มี closed-form solution แบบ Normal Eq 😢
- แต่สามารถหา gradient/partial derivative ได้ 😊

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^m (\sigma(\theta^T \cdot \mathbf{x}^{(i)}) - y^{(i)}) x_j^{(i)}$$

- ดังนั้น ใช้วิธี Stochastic Gradient Descent แบบเดิมได้

Lab: ใช้ Logistic Regression จำแนกพันธุ์ดอกไม้

```
from sklearn.linear_model import LogisticRegression
```



เป้าหมาย: จำแนก Iris-Virginica ออกจากชนิดอื่น โดยใช้ขนาดของกลีบ Sepal/Petal - width/length

การบ้าน Linear Regression

- from sklearn import datasets
data = datasets.load_boston()

Midterm

- AI
 - Strong vs Weak
 - Turing Test
- Machine Learning
 - เป้าหมายของ Machine learning
 - กระบวนการ training, testing/predicting
 - supervised vs unsupervised
 - classification vs regression
 - feature, class, model, evaluation metrics, bias-variance tradeoff, overfitting, cross-validation
- Supervised Learning
 - kNN (ข้อเสียคืออะไร, ค่า k ส่งผลกระทบ boundary อย่างไร)
 - Linear Regression (วิธี gradient descent ดีกว่า normal eq อย่างไร, regularize ทำเพื่ออะไร)
 - สมการ model, MSE cost function, normal equation, gradient descent, regularization
 - Logistic Regression (ใช้ทำอะไรได้, นิ๊กถึง iris)
 - สมการ model, sigmoid cost function