

Logistic Regression

อ. ปรัชญ์ ปิยะวงศ์วิศาล

Pratch Piyawongwisal

Today

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

Midterm Solution

Recap: Supervised Learning

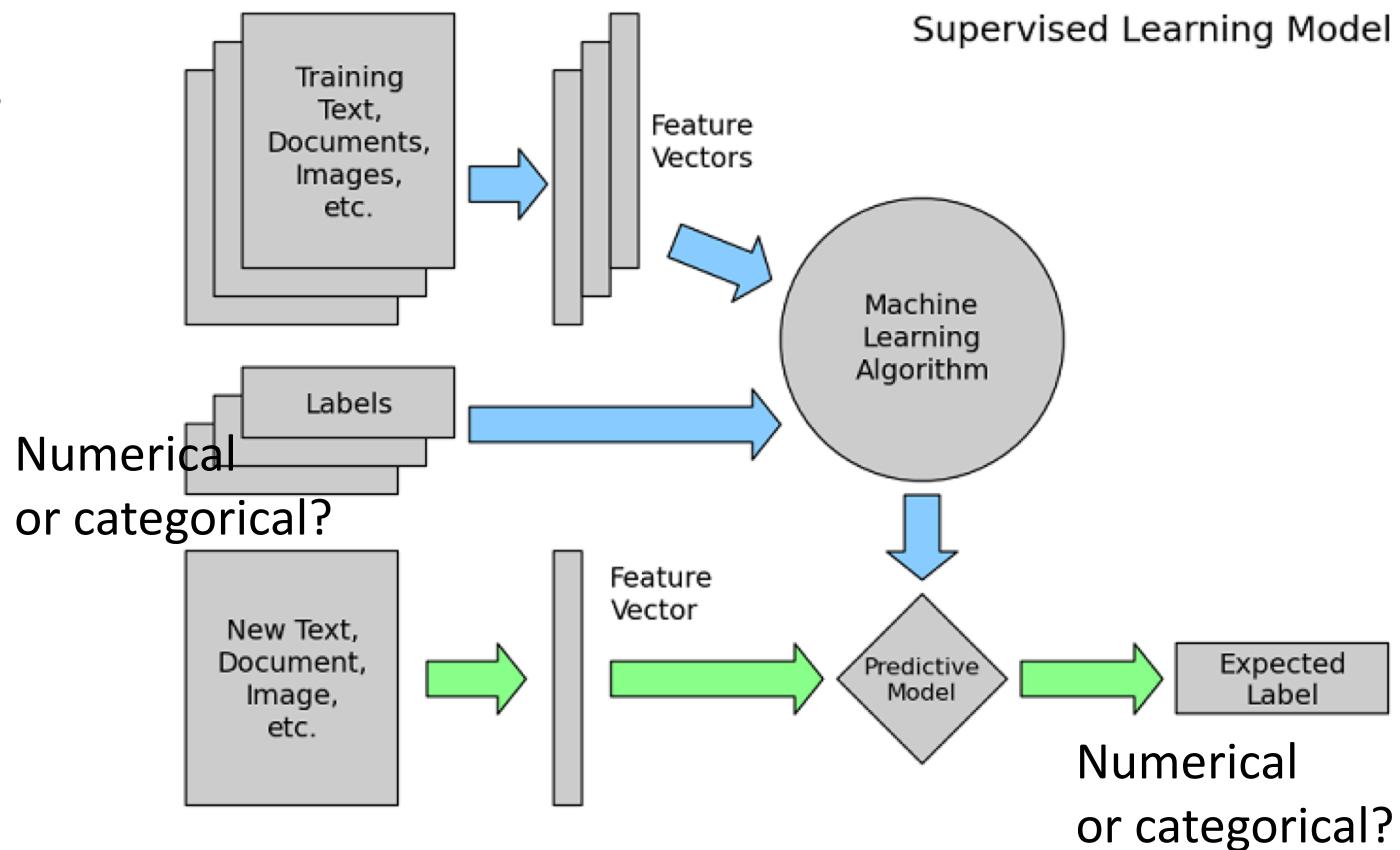
- Classification

kNN

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

- Regression

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



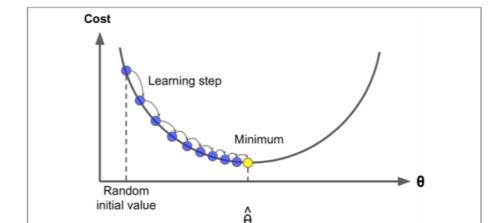
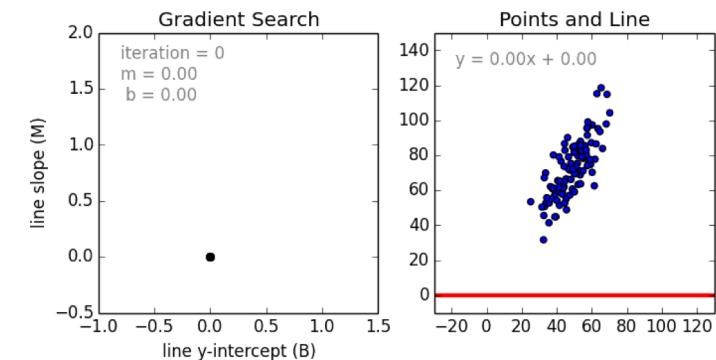
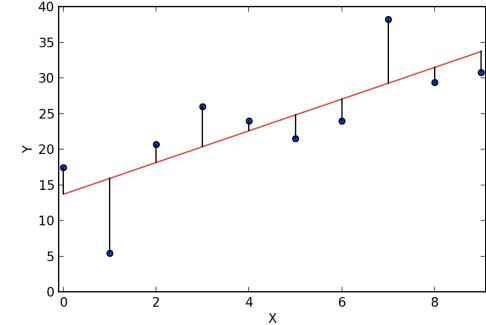
Linear Regression - Summary

- 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 :

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



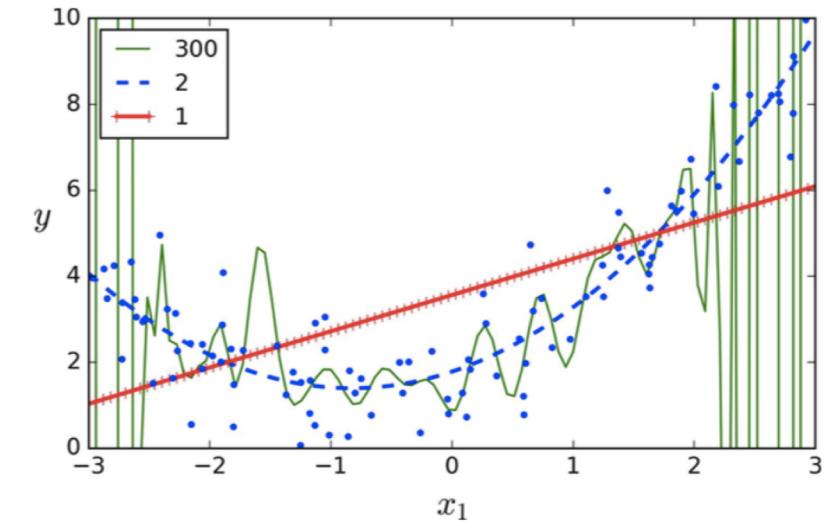
ภูเข้า MSE

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

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

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

Logistic Regression

How to apply?

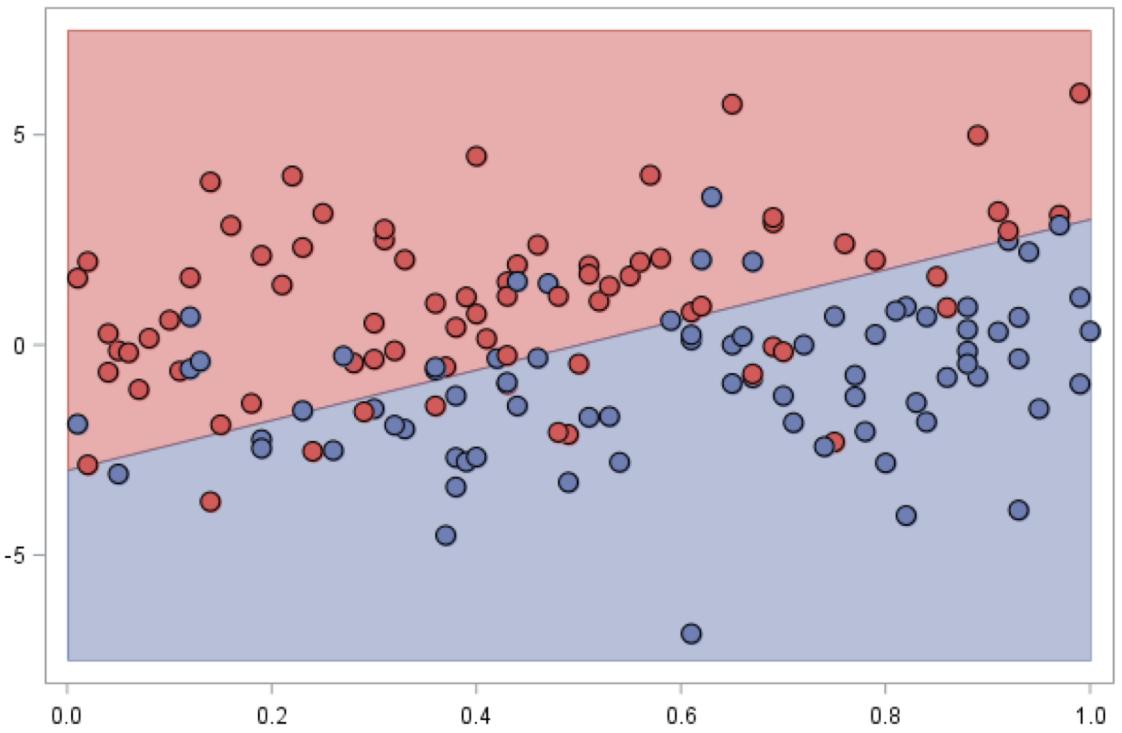
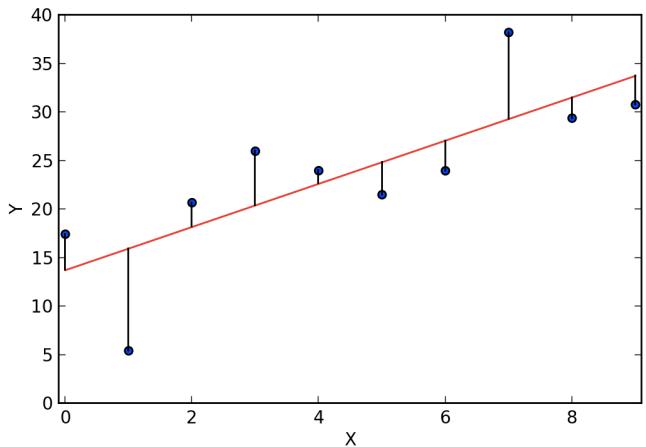
Linear Regression → (Binary) Classification

Linear Regression

- Input: x คือ input features
- Output: \hat{y} เป็นจำนวนจริง

โมเดล: $\hat{y} = h_{\theta}(x) = \theta^T x$

- ความสัมพันธ์แบบ linear



พิจารณาปัญหา binary classification เช่น

- อีเมลเป็น spam (label=1) หรือไม่เป็นสแปม (label=0)
- คนไข้เป็นมะเร็ง (label=1) หรือไม่เป็นมะเร็ง (label=0)

* \hat{p} คือค่าเป็นความน่าจะเป็น

Logistic Regression

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

Logistic Regression

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

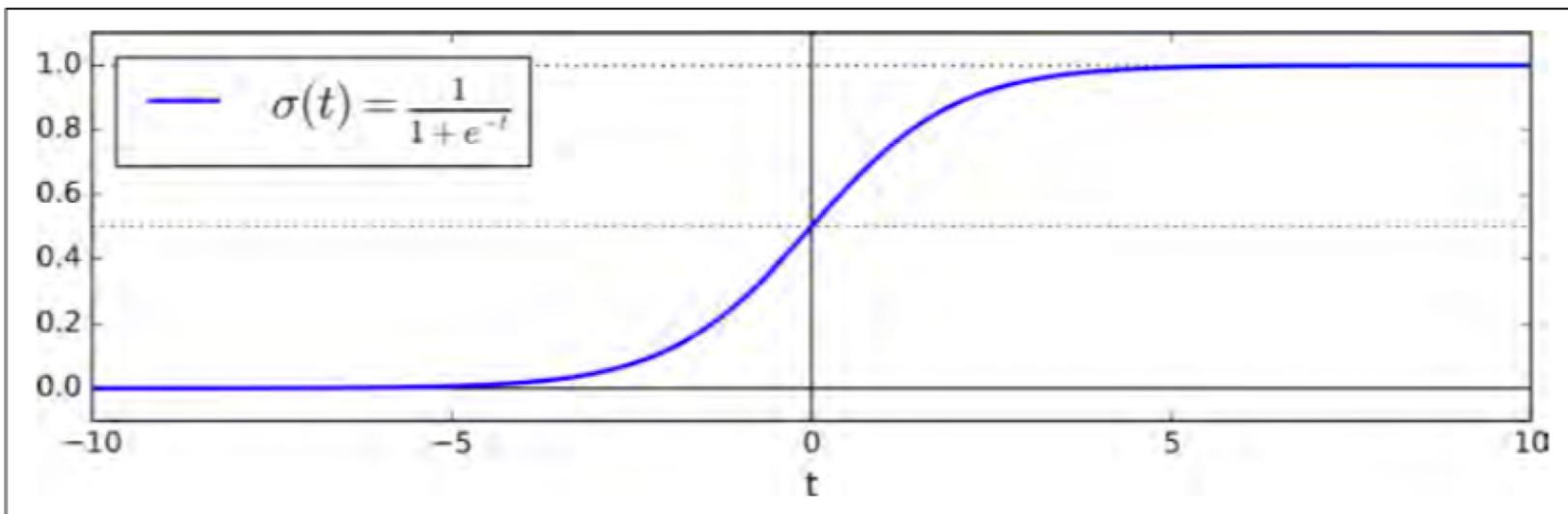
Sigmoid Function

- Solution: แค่เปลี่ยน model เป็น

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

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

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

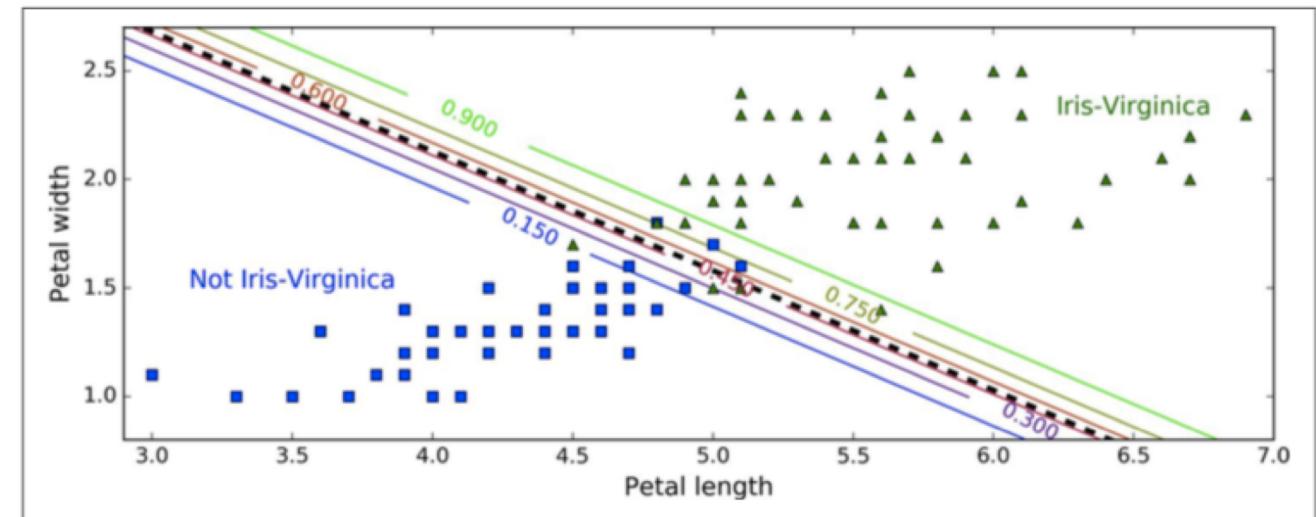


Logistic Regression Model

- โมเดล: $\hat{p}(y = 1|x; \theta) = h_{\theta}(x) = \sigma(\theta^T x)$
- กฎการตัดสินใจ

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

- โดยที่
 - Input: x คือ input features
 - Output: \hat{y} เป็น binary (0/1)

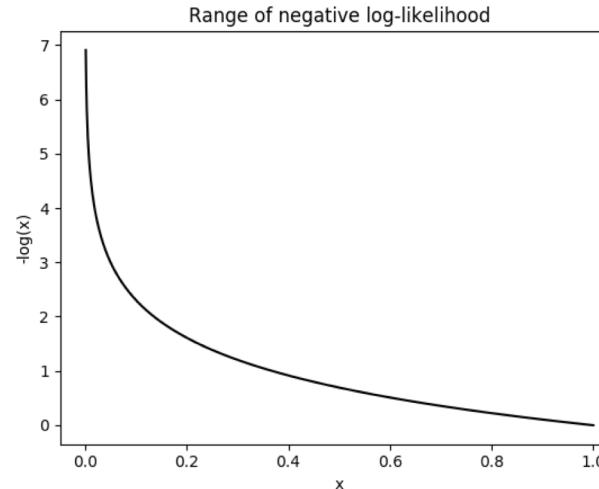


ตัวอย่าง **decision boundary** จากการใช้ชีวิต logistic regression

Cost Function

- 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

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

How to train?

- จะ train ไม่เดلنี้อย่างไร? (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 😞

How to train?

- จะ train โมเดลนี้อย่างไร? (minimize cost อย่างไร?)

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

- 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

From Binary to Multi-class classification

- ข้อจำกัดของ Logistic Regression คือ สามารถใช้กับข้อมูลที่มี label เพียง 2 คลาสเท่านั้น
- หากต้องการ generalize ไปใช้กับ multi-class สามารถเปลี่ยนไปใช้ softmax function

$$\hat{p}_k = \sigma(s(\mathbf{x}))_k = \frac{\exp(s_k(\mathbf{x}))}{\sum_{j=1}^K \exp(s_j(\mathbf{x}))}$$

- K is the number of classes.
- $s(\mathbf{x})$ is a vector containing the scores of each class for the instance \mathbf{x} .
- $\sigma(s(\mathbf{x}))_k$ is the estimated probability that the instance \mathbf{x} belongs to class k given the scores of each class for that instance.
- และ cost function จะเปลี่ยนเป็น Cross-Entropy ซึ่งจะเหมือน LR cost หาก $k=2$

$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(\hat{p}_k^{(i)})$$

Lab: ใช้ Logistic Regression จำแนกพันธุ์ดอกกล้วยไม้

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



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