Quiz I

66104 02230
ศิริสุข ทานธีรธรรม

$$D = \{ (x_i^{1}, x_i^{2}, x_i^{3}, \dots, y_i), \dots \}$$

feature 1, feature 2, feature 3, feature (d)

$$\underbrace{X}_{\text{feature space}} \in \mathbb{R}^{d} \qquad x_i = \left.\begin{bmatrix} \text{age} \\ \text{position} \\ \text{experience} \\ \vdots \end{bmatrix}\right\} d$$

ตอบ $\underbrace{Y}_{\text{label space}} \in \{\text{salary}\}$

1. If you were to do supervised learning to make predictions about your start salary.

ตอบ เนื่องจากเป็นการ learning เพื่อ predict ค่าเงินเดือน linear

1.1. What type of supervised learning are you in? จึงอยู่ในประเภท regression

1.2. Declare how you will define your data set D with feature space X and label space Y. Also, give an example to illustrate a sample in D.

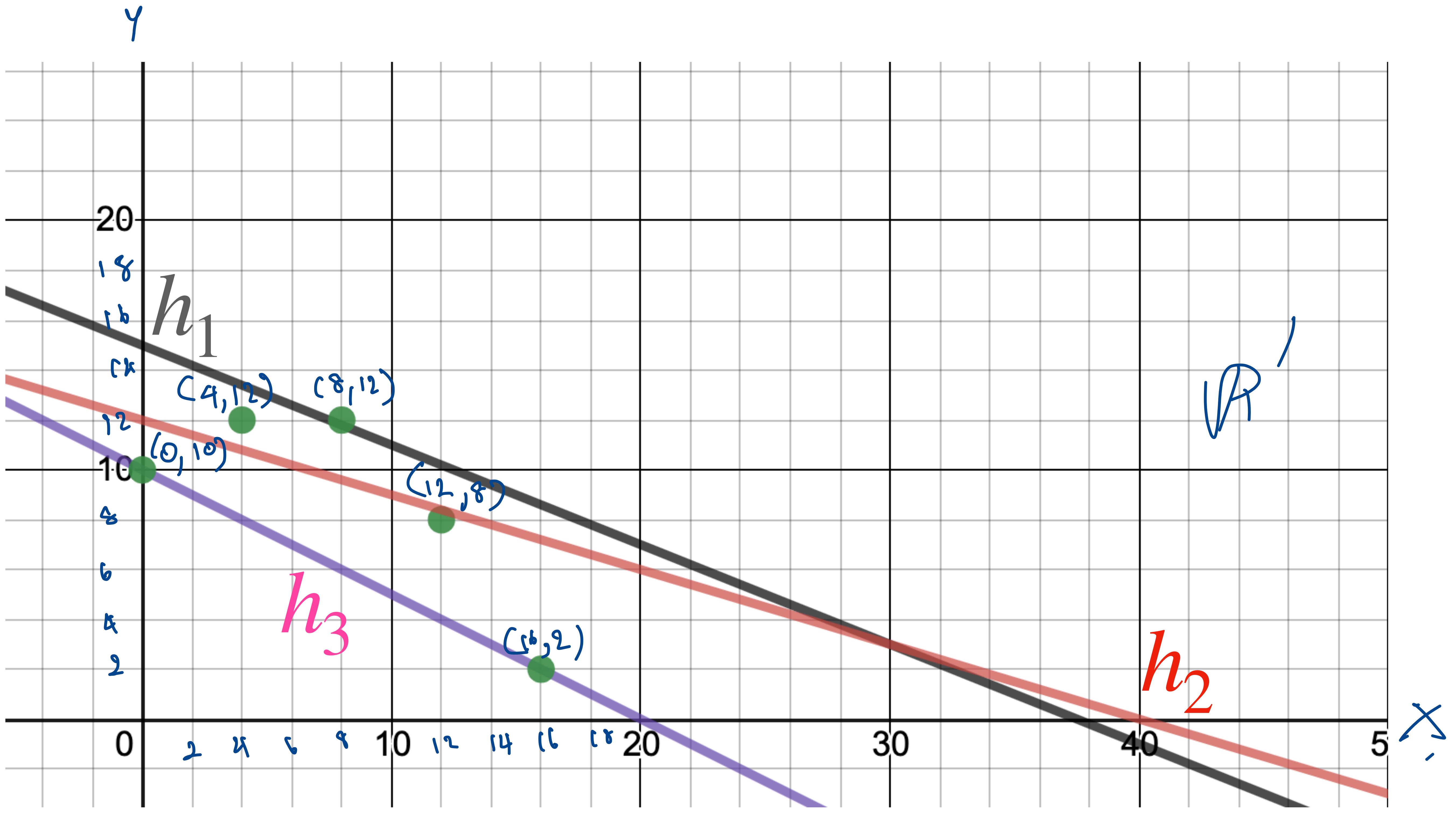
1.3. Describe how would you collect the data set D?

ตอบ ข้อมูลจาก kaggle, database, survey, ask for data

6610402230 ศิริกุล ทานเจริญ

Quiz I (cont.)

2. Suppose we have 5 data points of the set D the hypothesis class $\mathscr{H} = \{h_1, h_2, h_3\}$ in the following plot.



6610402230 ศิริสุข ทานธรรม

Quiz I (cont.)

$|D| = 5$

$l_{0/1}(h_1, D) = \frac{4}{5}$

$l_{0/1}(h_2, D) = \frac{4}{5}$

$l_{0/1}(h_3, D) = \frac{3}{5}$

ตอบ
เลือก h_3 เพราะ
จาก $l_{0/1}(h, D)$ จาก
ทุก Hypothesis class $h \in H$
h_3 ให้ค่า loss น้อยที่สุด ถ้าไม่
แต่อาจไม่เหมาะมาก เพราะ $l_{0/1}$ loss เหมาะกับ Regression

2.1.What is the best function in $\mathscr{H}$ if we are to use 0/1 loss function to measure the losses? Please demonstrate your solution.

$$\mathscr{L}_{0/1}(h, D) = \frac{1}{|D|} \sum_{\forall(\overrightarrow{x},y)\in D} \delta(h(\overrightarrow{x}), y), \text{ where } d(h(\overrightarrow{x}), y) = \begin{cases} 1, & \text{if } h(\overrightarrow{x}) \neq y \\ 0, & \text{otherwise} \end{cases}$$

2.2.What is the best function in $\mathscr{H}$ if we are to use square loss function $\mathscr{L}_{sq}$ to measure the losses? Please demonstrate your solution.

ตอบ h_2
ที่สุด

เพราะ

$$\mathscr{L}_{sq}(h, D) = \frac{1}{|D|} \sum_{\forall(\overrightarrow{x},y)\in D} (h(\overrightarrow{x}) - y)^2$$

จาก เส้น $h_{1,2,3}$ ถ้าถูกหา l_{sq} จะได้ว่า h ตัวอื่นเมื่อคิด $\Sigma(\hat{y}-y)^2$ มีเส้นลากผ่านที่ไกล data point กว่า
ทำไมเพื่อ หาร $|D|$ (กรณี) loss จะมีค่ามากกว่า h_2 และ h_2 มี ความ generalize กว่า (dist ไม่เกาะ data point เดิม ไป)