Classification

Classification and Representation

Logistic Regression

Classification

- Email: Spam / Not Spam?
- Online Transactions: Fraudulent (Yes / No)?
- Tumor: Malignant / Benign ?

Classification

- Email: Spam / Not Spam?
- Online Transactions: Fraudulent (Yes / No)?
- Tumor: Malignant / Benign ?

$y \in \{0, 1\}$

0: “Negative Class” (e.g., benign tumor)

1: “Positive Class” (e.g., malignant tumor)

Classification

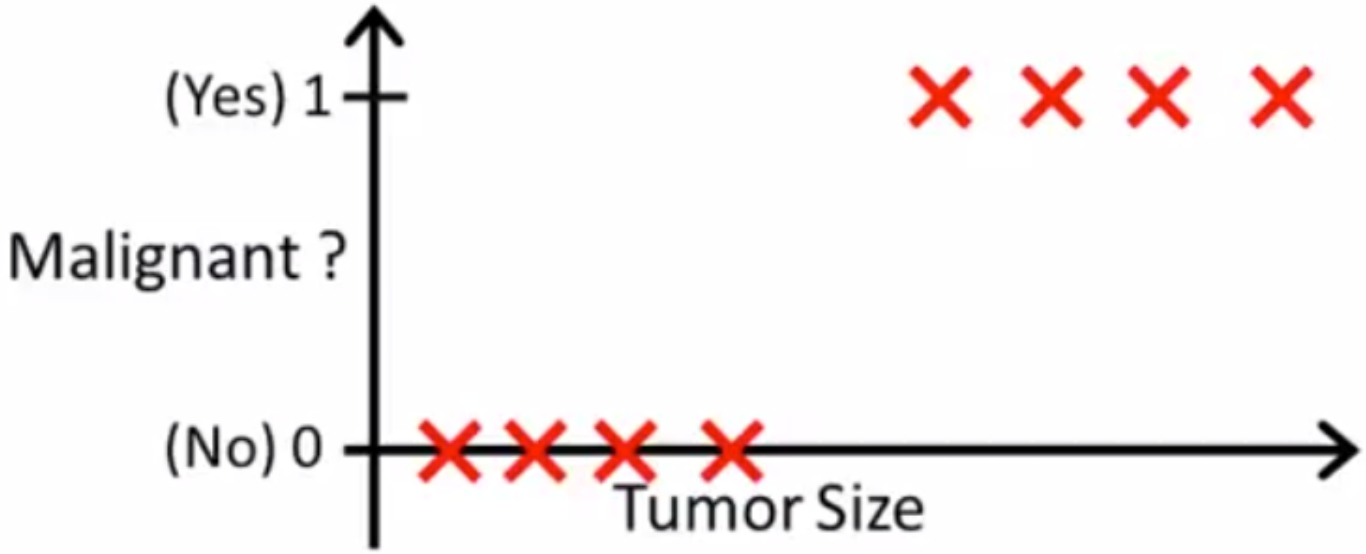
- Email: Spam / Not Spam?
- Online Transactions: Fraudulent (Yes / No)?
- Tumor: Malignant / Benign ?

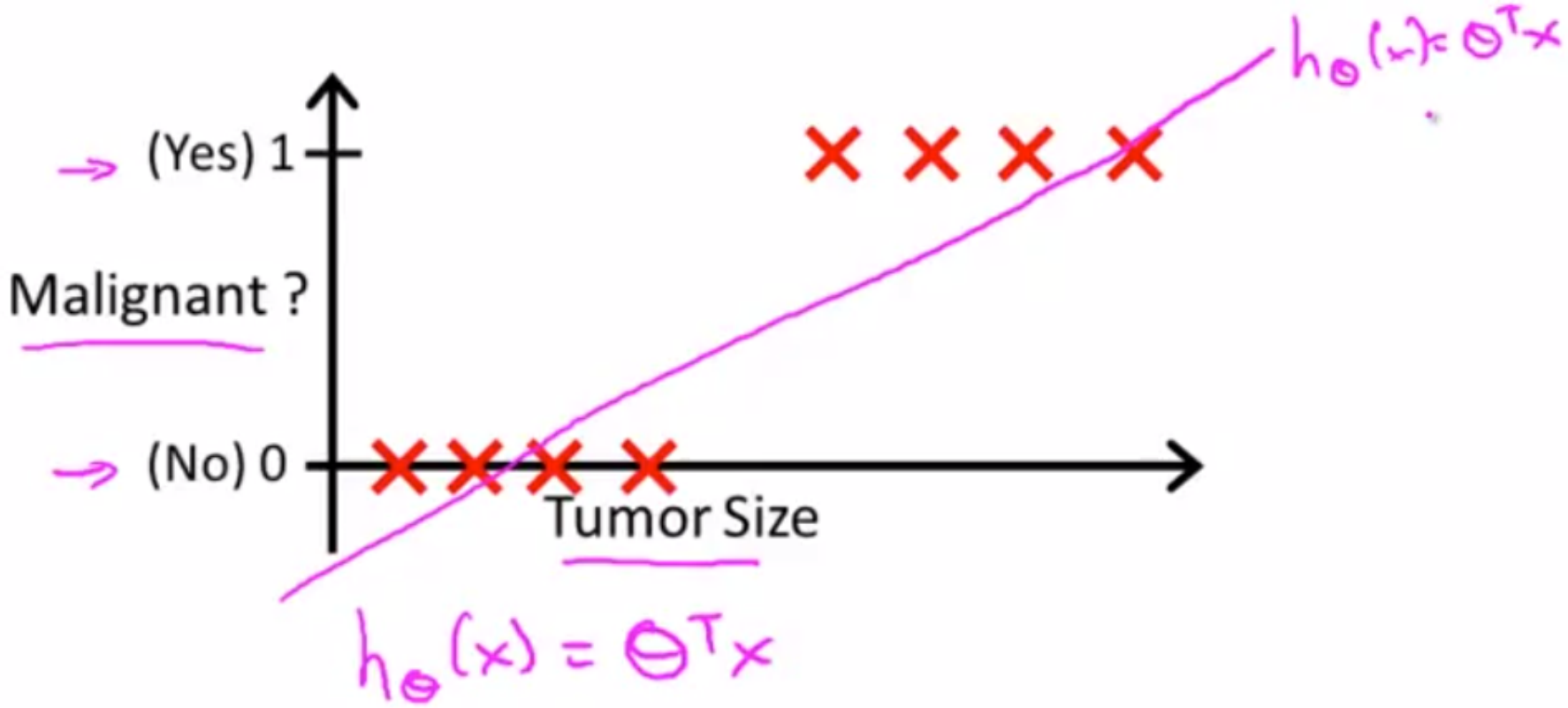
$y \in \{0, 1\}$

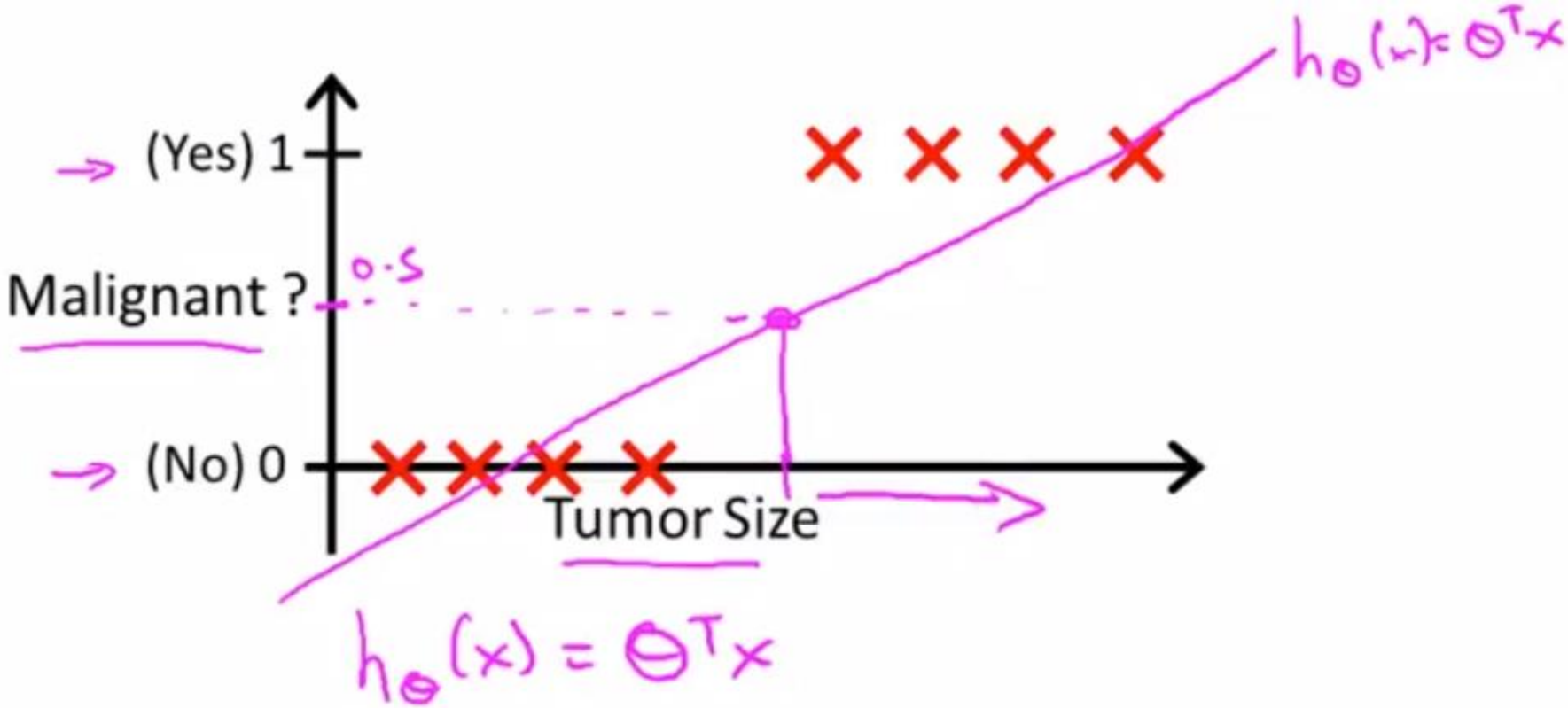
0: "Negative Class" (e.g., benign tumor)

1: "Positive Class" (e.g., malignant tumor)

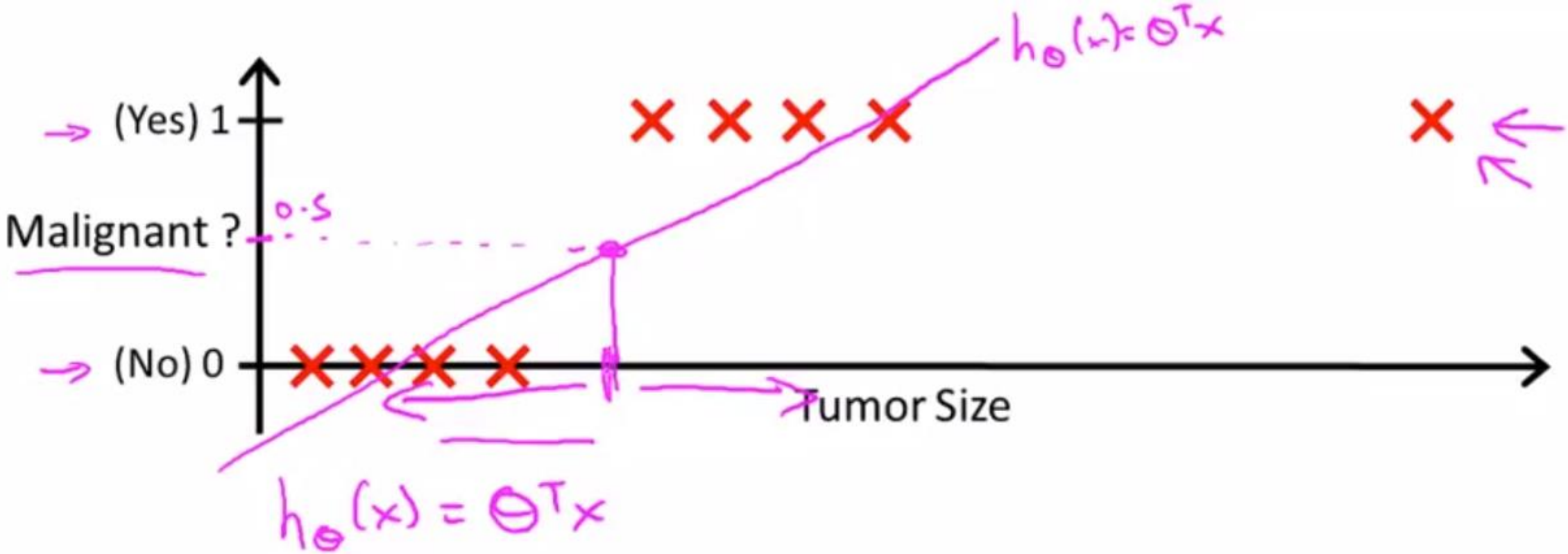
$y \in \{0, 1, 2, 3\}$



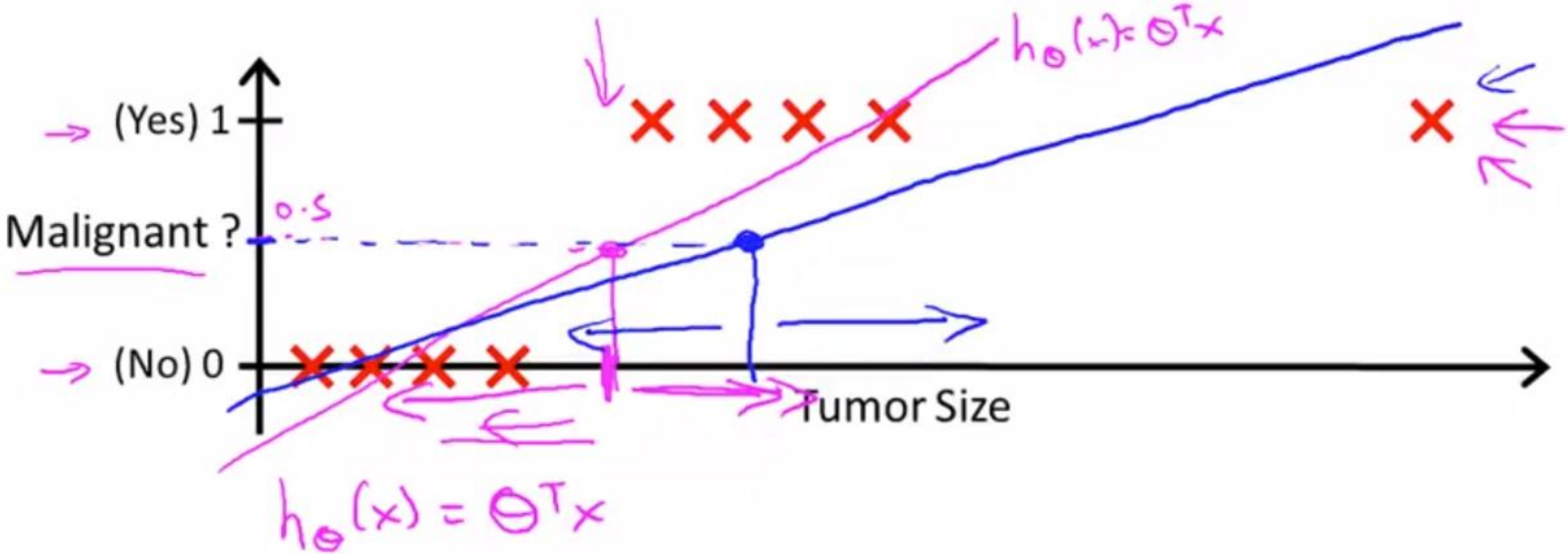




- Threshold classifier output $h_\theta(x)$ at 0.5:
  - If $h_\theta(x) \geq 0.5$, predict "y = 1"
  - If $h_\theta(x) < 0.5$, predict "y = 0"



- Threshold classifier output $h_\theta(x)$ at 0.5:
  - If $h_\theta(x) \geq 0.5$, predict "y = 1"
  - If $h_\theta(x) < 0.5$, predict "y = 0"



- Threshold classifier output $h_\theta(x)$ at 0.5:
  - If $h_\theta(x) \geq 0.5$, predict "y = 1"
  - If $h_\theta(x) < 0.5$, predict "y = 0"

Classification: y = 0 or 1

$h_\theta(x)$ can be > 1 or < 0

Classification: y = 0 or 1

$h_\theta(x)$ can be > 1 or < 0

Exercise

Which of the following statements is true?

- If linear regression doesn't work on a classification task as in the previous example, applying feature scaling may help.
- If the training set satisfies $0 \leq y(i) \leq 1$ for every training example $(x(i), y(i))$, then linear regression's prediction will also satisfy $0 \leq h(x) \leq 1$ for all values of x.
- If there is a feature x that perfectly predicts y, i.e if $y=1$ when $x \geq c$ and $y=0$ whenever $x<c$ (for some constant c), then linear regression will obtain zero classification error.
- None of the above statements are true.

Classification: y = 0 or 1

$h_\theta(x)$ can be > 1 or < 0

Logistic Regression: $0 \leq h_\theta(x) \leq 1$

Classification: $y = 0$ or 1

$h_\theta(x)$ can be > 1 or < 0

Logistic Regression: $0 \leq h_\theta(x) \leq 1$

↳ Classification