

# Machine Learning

For Beginners

Dr. Chetan Jaiswal

Computer Science

Truman State University

# What is machine learning?

- "A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ ."

Example: playing checkers.

$E$  = the experience of playing many games of checkers

$T$  = the task of playing checkers.

$P$  = the probability that the program will win the next game.

“A computer program is said to *learn* from experience  $E$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$ .”

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task  $T$  in this setting?

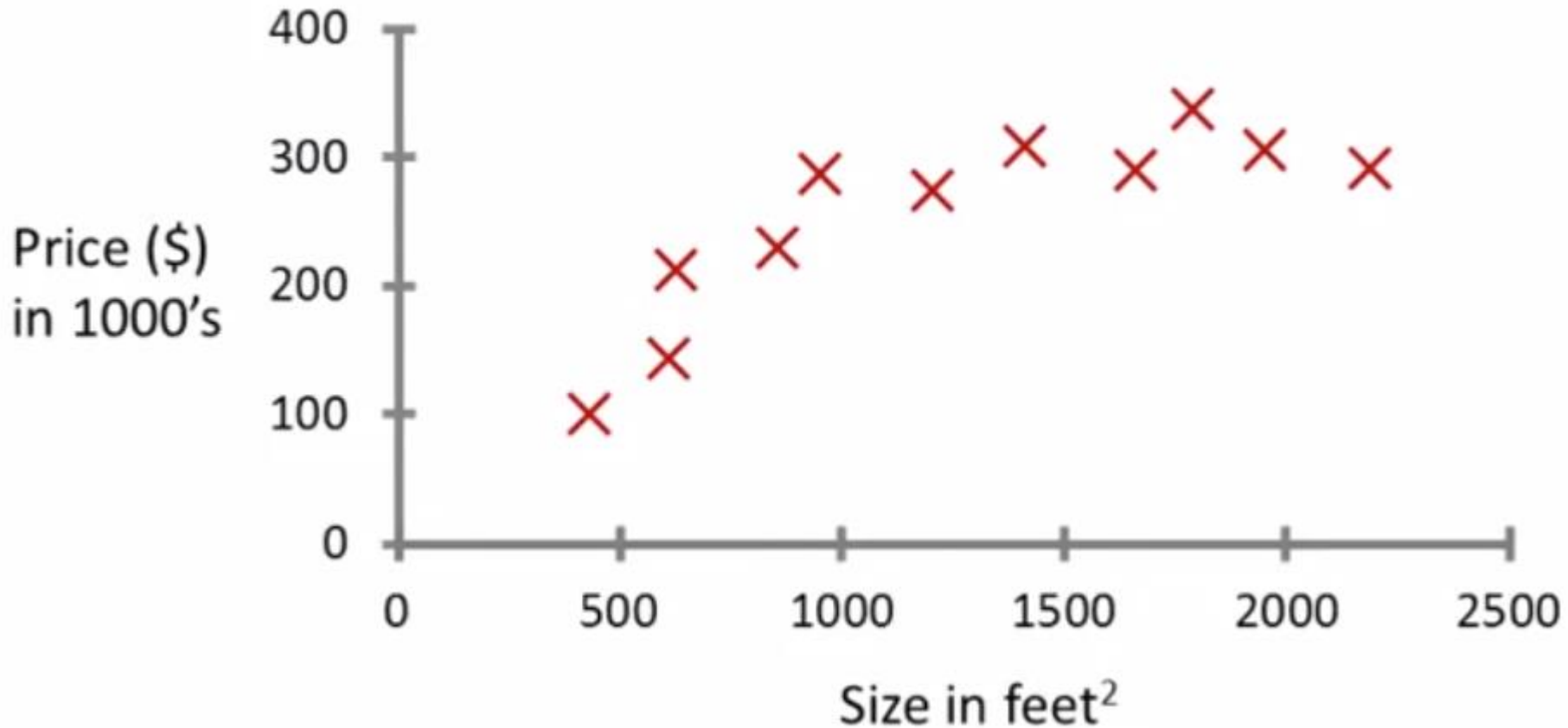
- ☐ Classifying emails as spam or not spam.
- ☐ Watching you label emails as spam or not spam.
- ☐ The number (or fraction) of emails correctly classified as spam/not spam.
- ☐ None of the above—this is not a machine learning problem.

# Supervised learning?

- In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.
- Supervised learning problems are categorized into "regression" and "classification" problems. In a regression problem, we are trying to predict results within a continuous output, meaning that we are trying to map input variables to some continuous function. In a classification problem, we are instead trying to predict results in a discrete output. In other words, we are trying to map input variables into discrete categories.

# Supervised learning?

## Housing price prediction.



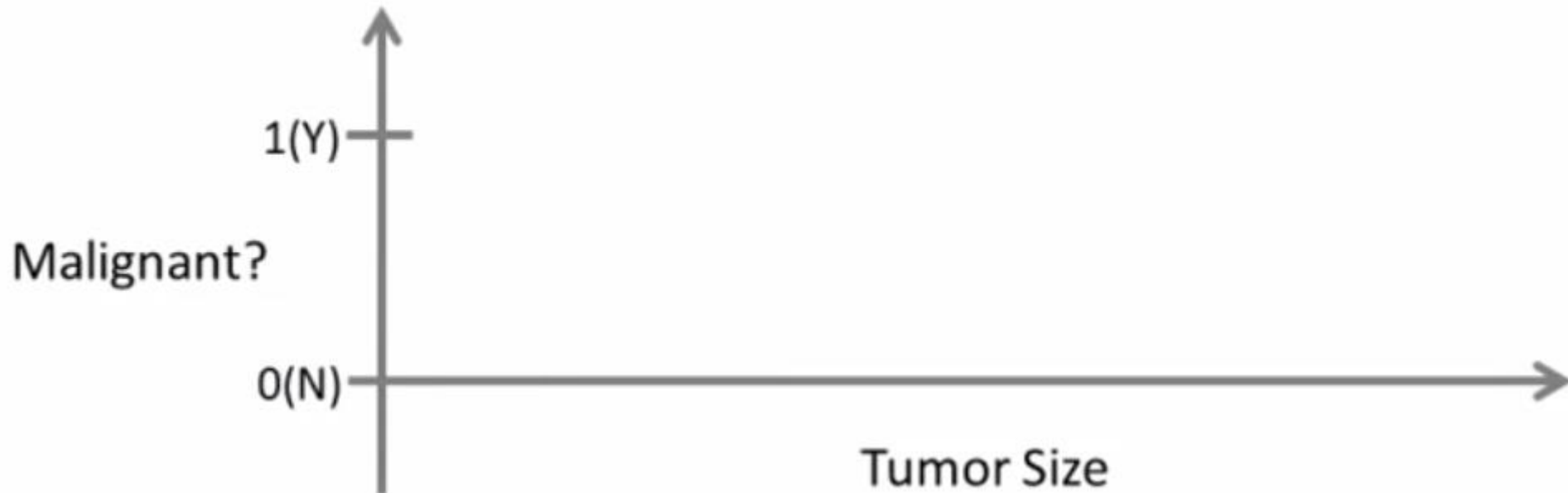
# Supervised learning?

- Right answers are given, the algorithm predicts “more” right answers
- Referred as regression

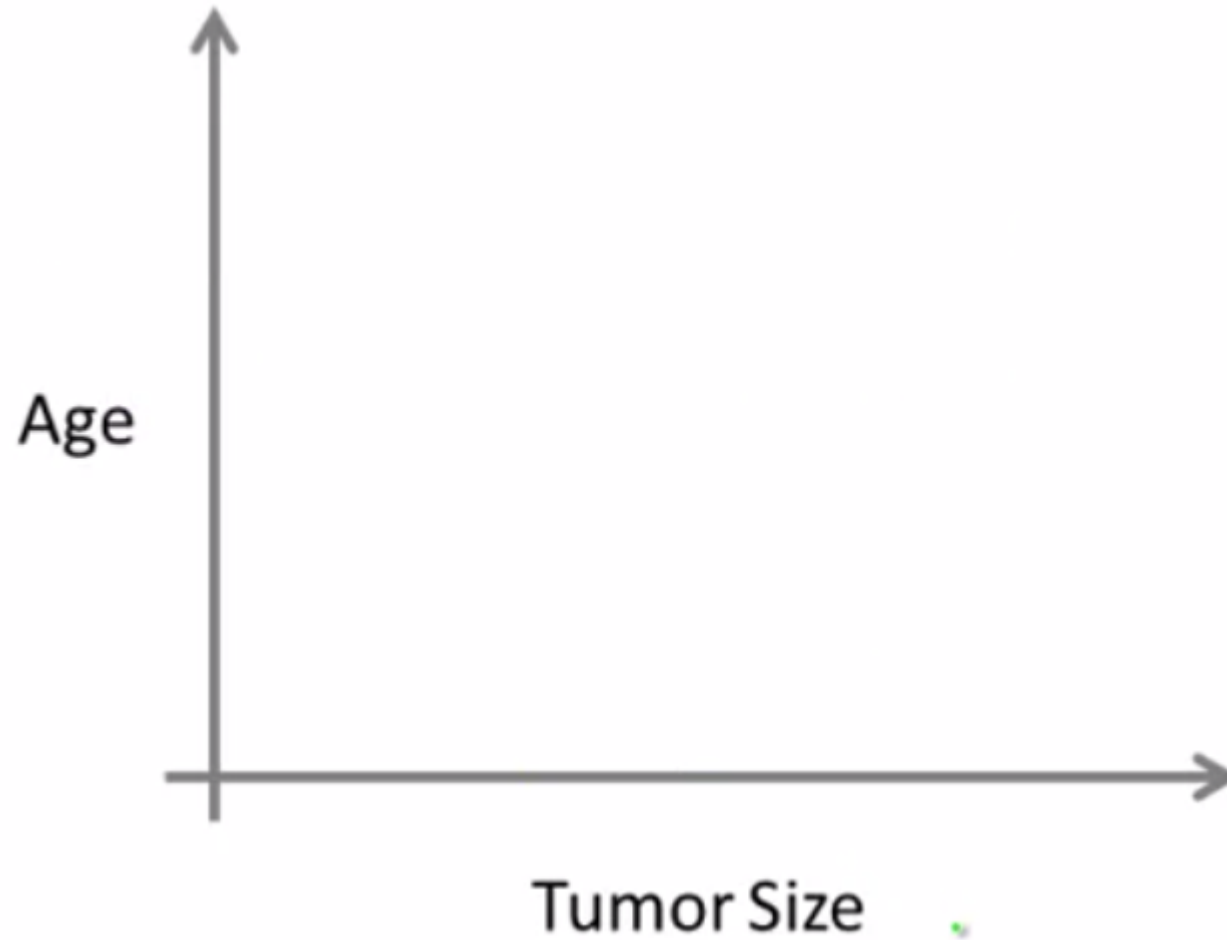
# Supervised learning?

Press Esc to exit full screen

## Breast cancer (malignant, benign)



# Supervised learning?





# Supervised learning?

You're running a company, and you want to develop learning algorithms to address each of two problems.

Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

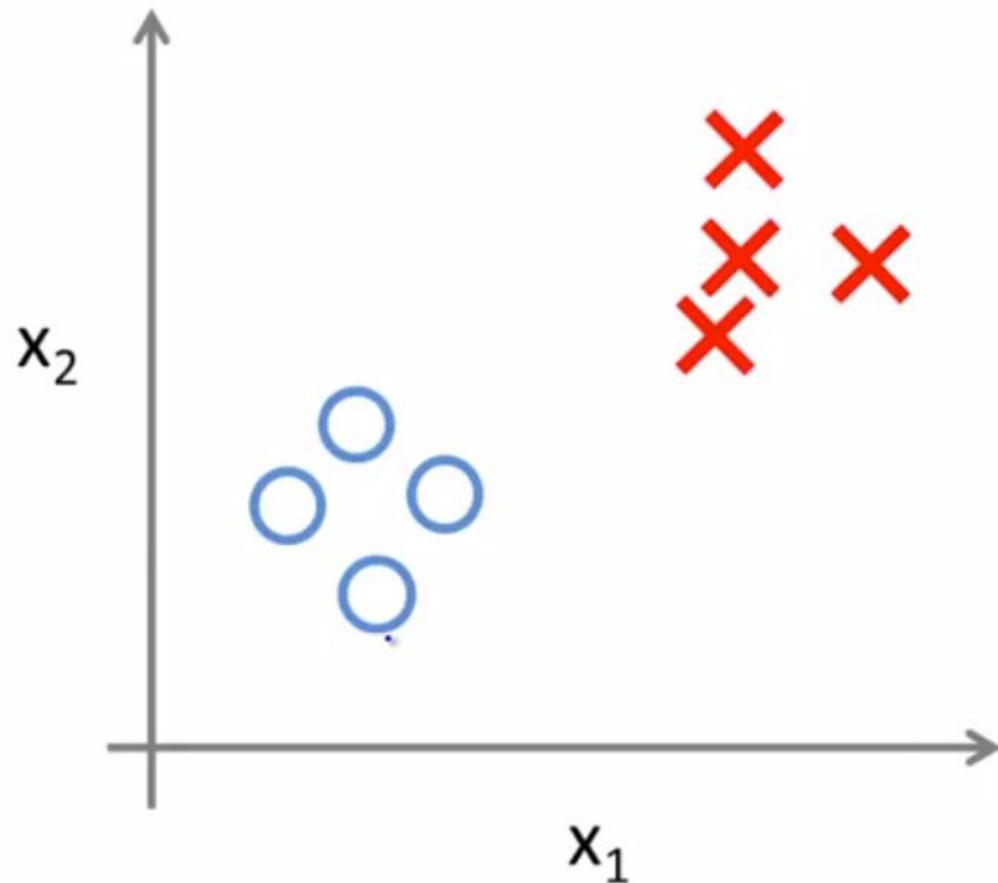
Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

Should you treat these as classification or as regression problems?

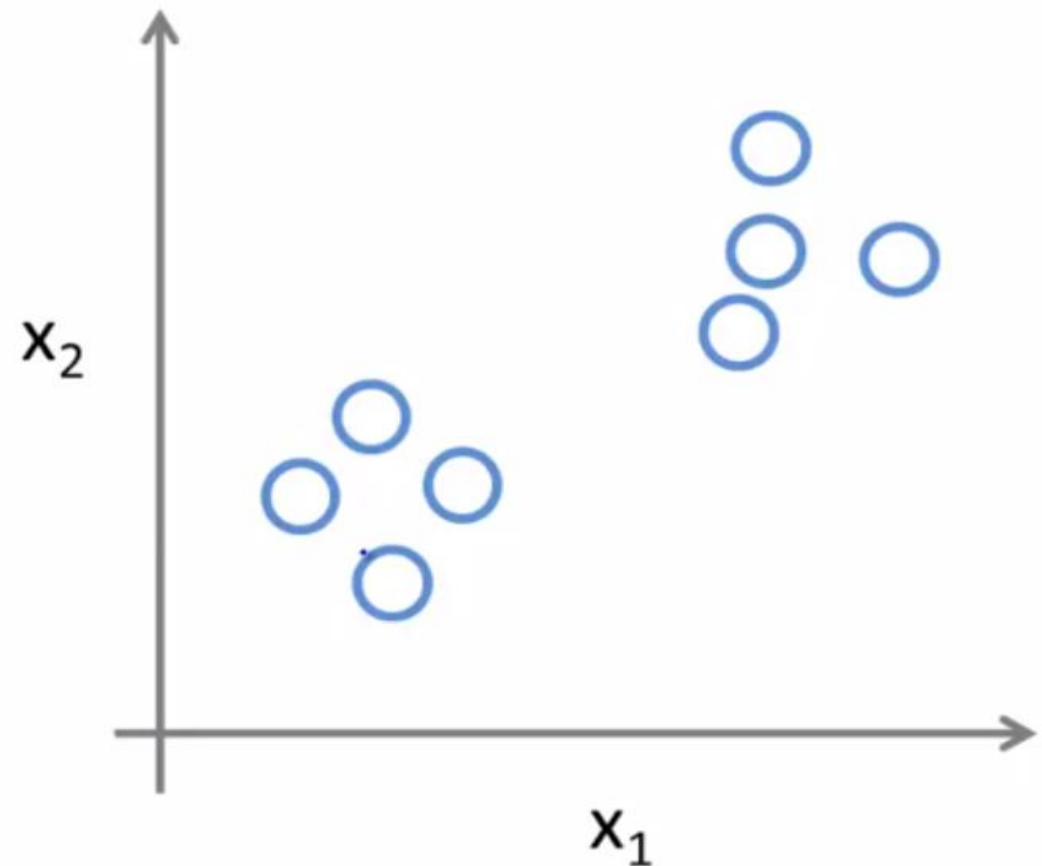
- ☐ Treat both as classification problems.
- ☐ Treat problem 1 as a classification problem, problem 2 as a regression problem.
- ☐ Treat problem 1 as a regression problem, problem 2 as a classification problem.
- ☐ Treat both as regression problems.

# Unsupervised learning?

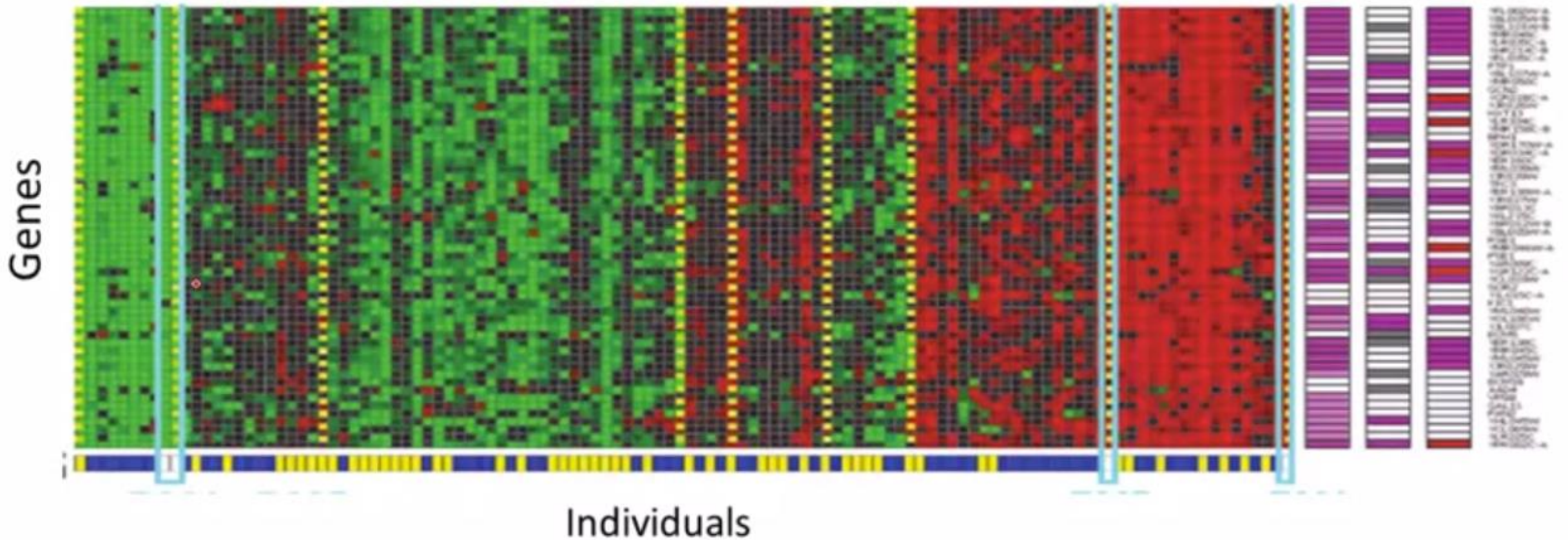
Supervised Learning



Unsupervised Learning



# Unsupervised learning?

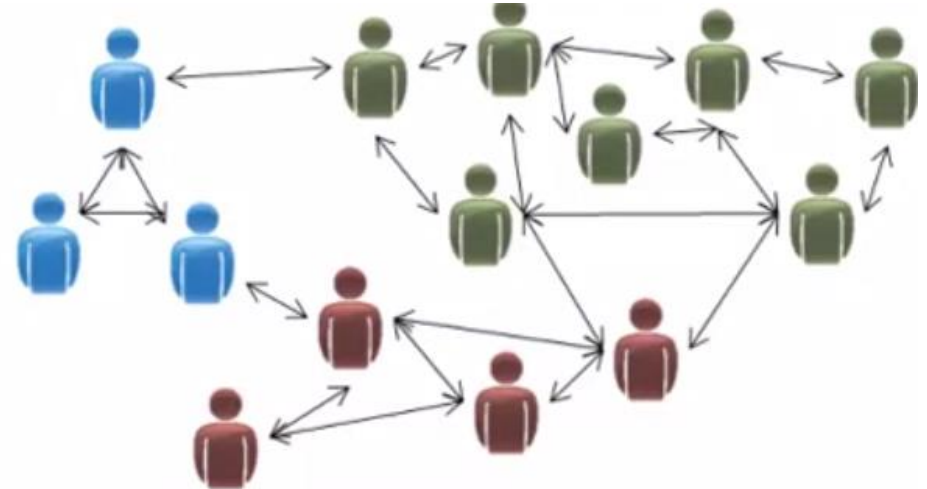




# Unsupervised learning?



Organize computing clusters



Social network analysis



Market segmentation



Astronomical data analysis

# Unsupervised learning?

Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

- ☐ Given email labeled as spam/not spam, learn a spam filter.
- ☐ Given a set of news articles found on the web, group them into set of articles about the same story.
- ☐ Given a database of customer data, automatically discover market segments and group customers into different market segments.
- ☐ Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.