An Introduction to Machine Learning

Northrup Grumman

STEM Camp (2017)

Game Time!

Alaskan Malamute or Siberian Husky?

Training Set

What distinguishing **features** do you notice?

Malamute

Husky

Testing Set

How can we do better?

How can we do better?

- **More training data**
  - More data   >   Better algorithm
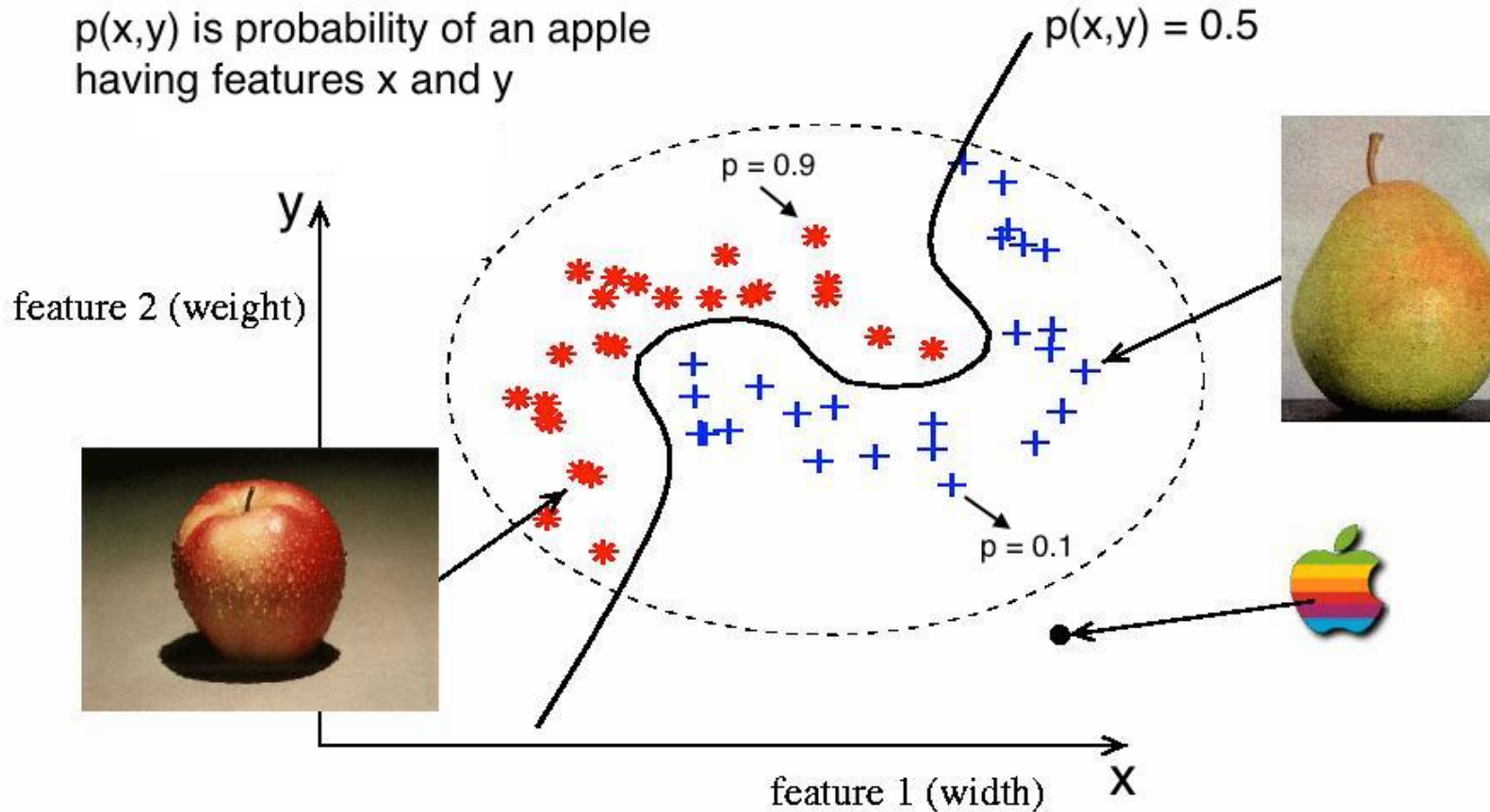
How can we do better?

- **More training data**
  - More data   >   Better algorithm

- Better picture quality
  - Clean data   >   More data

How can we do better?

- **More training data**
  - More data > Better algorithm

- Better picture quality
  - Clean data > More data

- More information about huskies and malamutes
  - Data is useless if we're looking in the wrong direction!
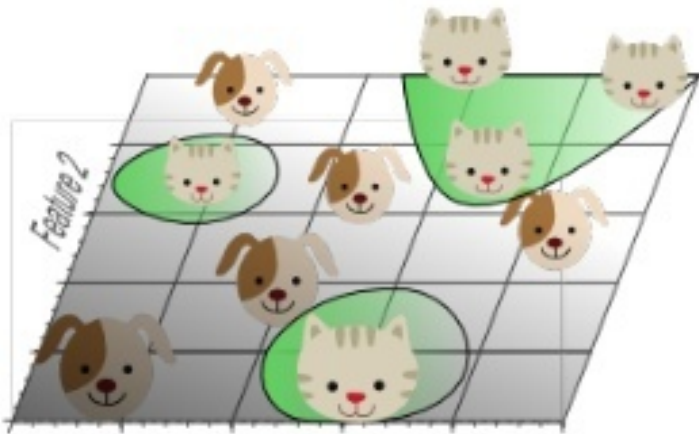
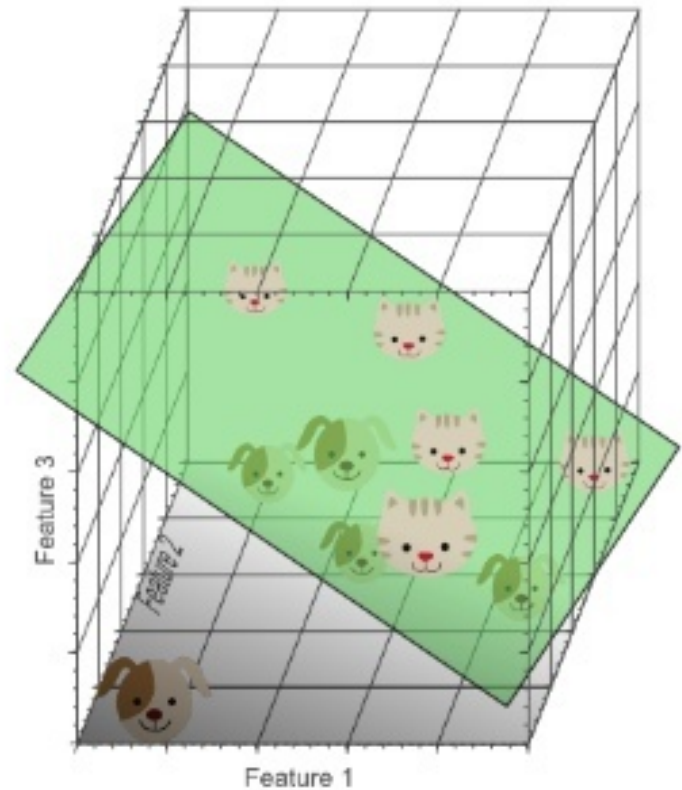# How do we make computers think?

Classification

p(x,y) is probability of an apple having features x and y

p(x,y) = 0.5

p = 0.9

p = 0.1

y

feature 2 (weight)

feature 1 (width)   x

Classification

1 feature

2 features

3 features

Classification

What the computer sees

image classification →
82% cat
15% dog
2% hat
1% mug

Supervised Learning
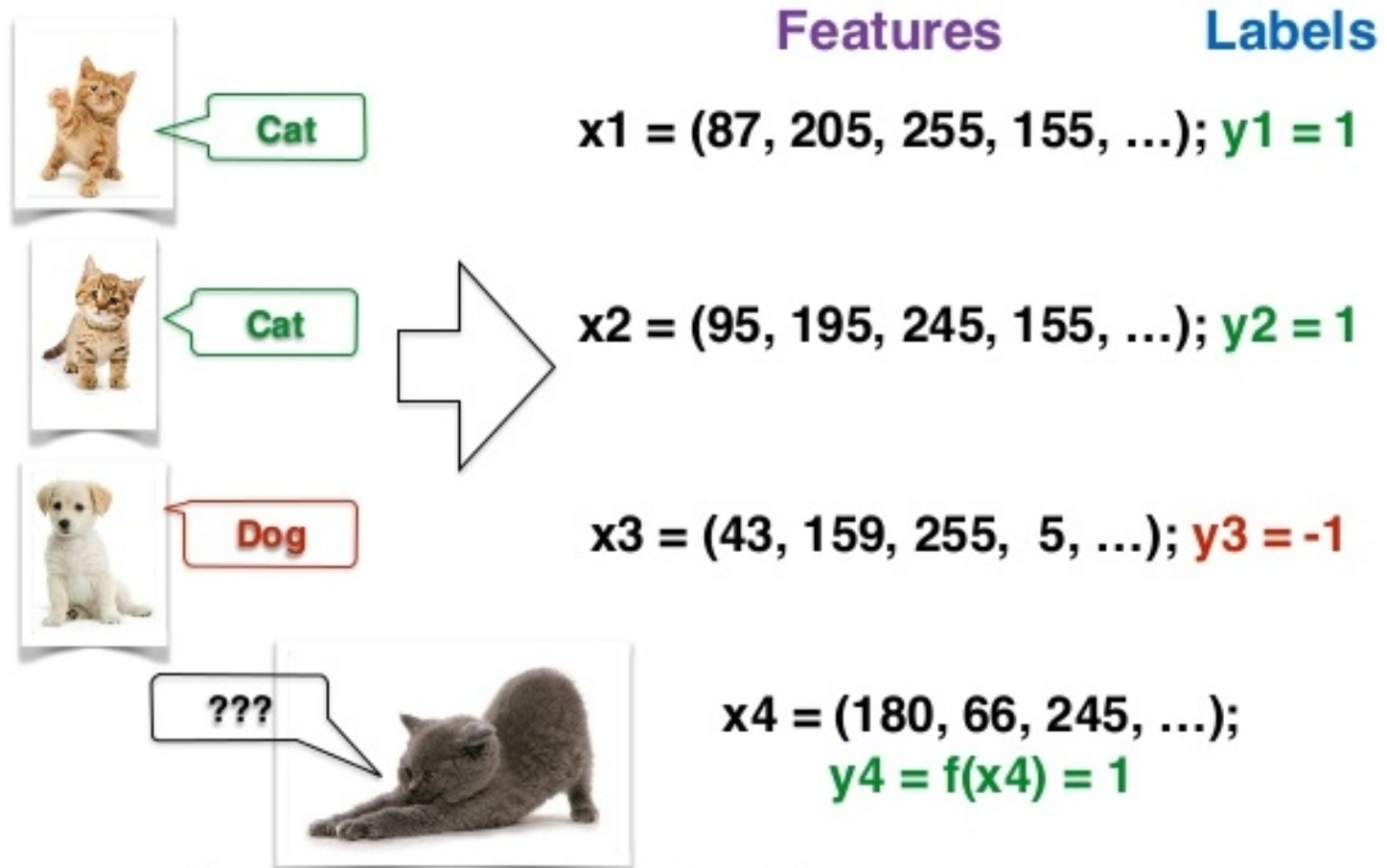
Features   Labels

$x_1 = (87, 205, 255, 155, \ldots); y_1 = 1$

$x_2 = (95, 195, 245, 155, \ldots); y_2 = 1$

$x_3 = (43, 159, 255, 5, \ldots); y_3 = -1$

$x_4 = (180, 66, 245, \ldots);$
$y_4 = f(x_4) = 1$

Supervised Learning

- The computer "sees" a training set,

- "Learns" the probability that a combination of features matches each class,

- "Predicts" a class for a new set of features

What to take away:

- **Probability** is the language of machine learning!

- **Data** quality and quantity are vital

- Computers don't see the way we do, and they **can't explain their decisions**

Coming Up:

- A lab on probability

- A lesson on Naïve Bayes, our first machine-learning algorithm

- Building your own Bayes-Law simulator