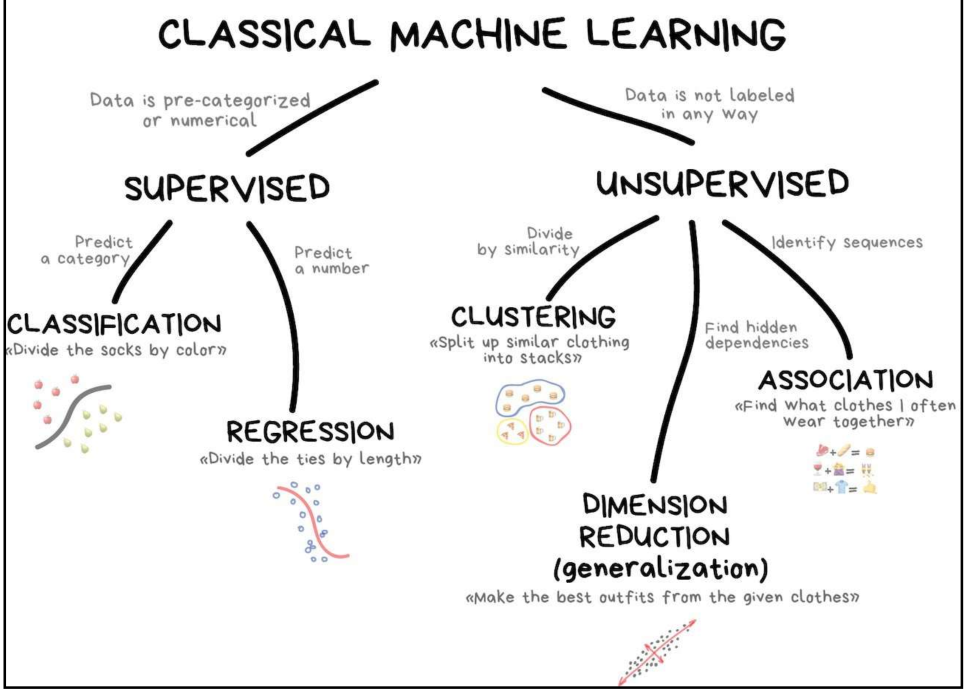
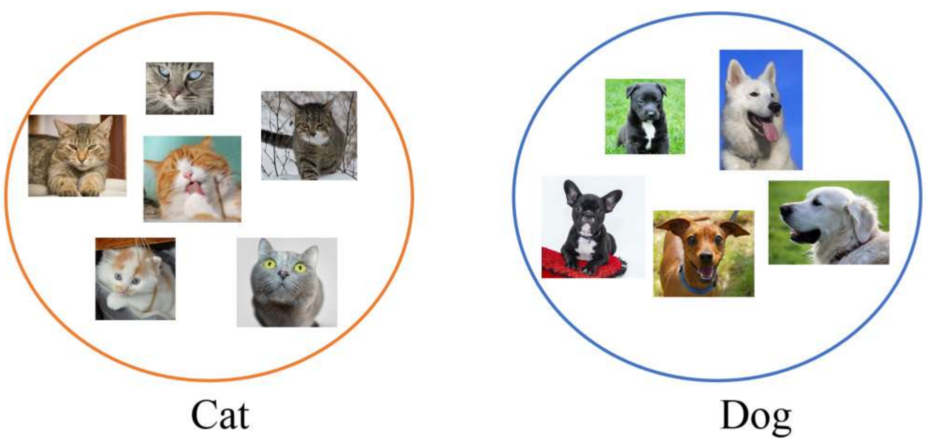
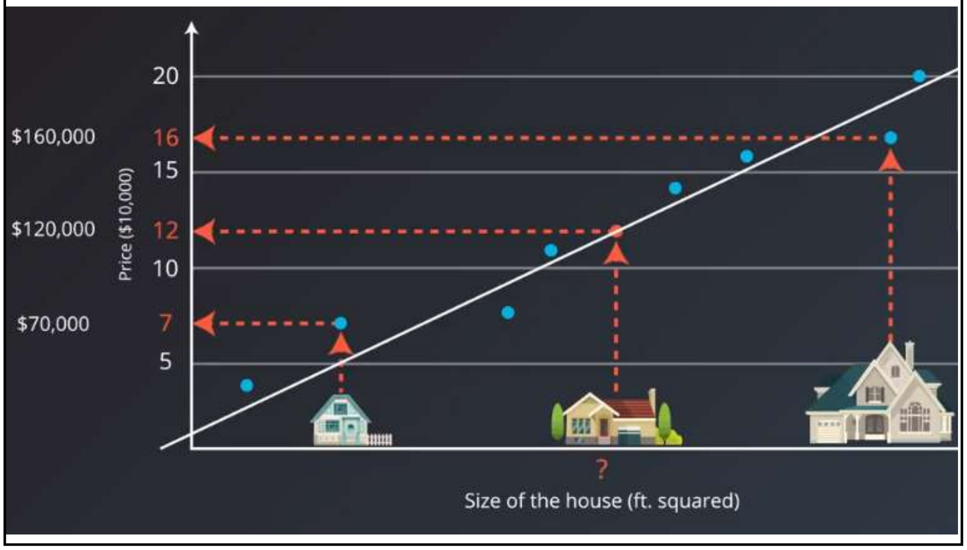


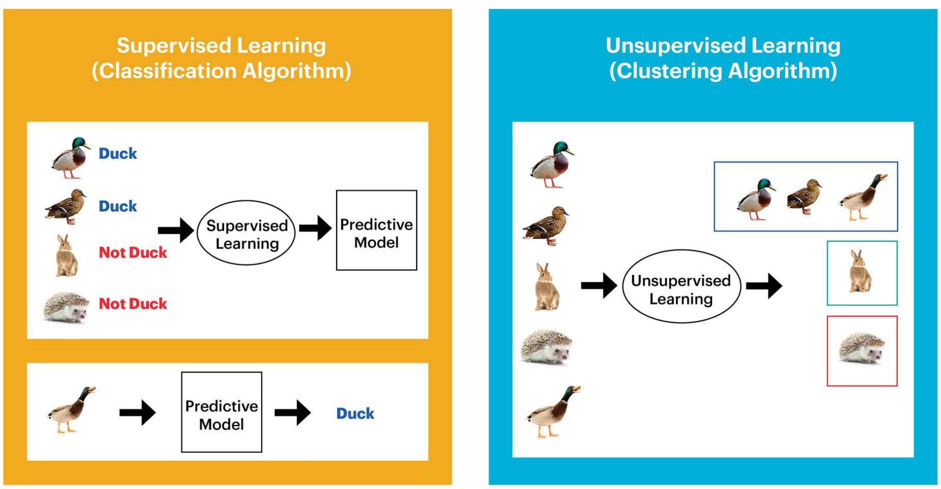
SUPERVISED/ CLASSIFICATION



SUPERVISED/ REGRESSION



UNSUPERVISED/ CLUSTERING



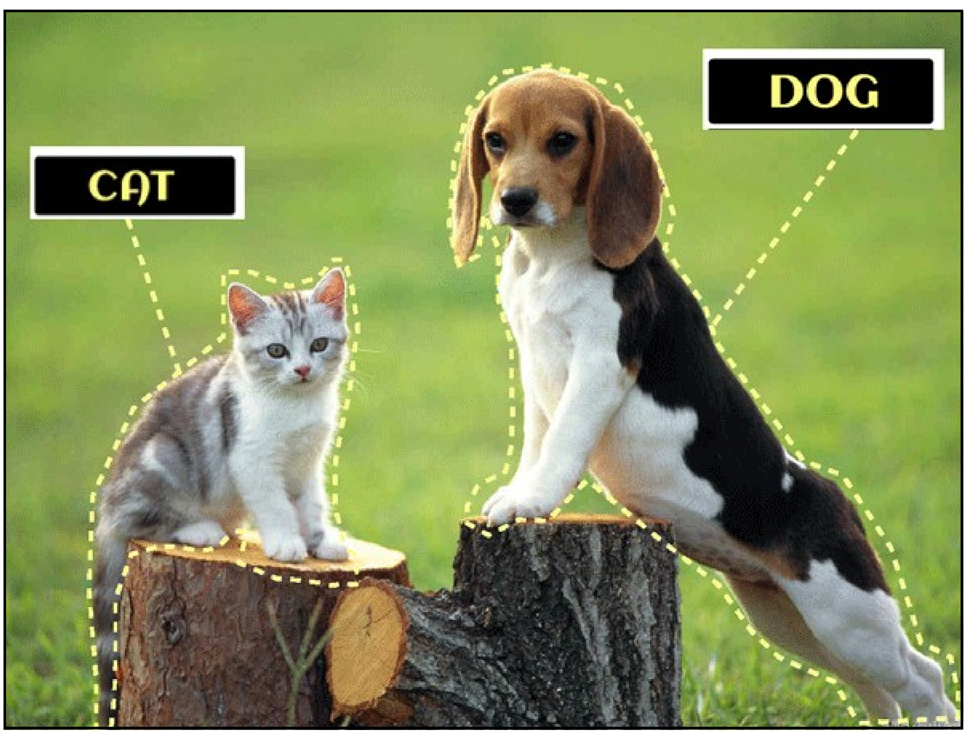
Machine Learning

- In the past decade, ML given us
  - Self-driving cars
  - Speech recognition
  - Real-time computer vision
  - Effective web search
  - Understanding human genome

Deep Learning?

1. A class of machine learning techniques that exploit many layers of non-linear information processing for supervised or unsupervised feature extraction and transformation, and for pattern analysis and classification.
2. A sub-field within machine learning that is based on algorithms for learning multiple levels of representation in order to model complex relationships among data.


Manfield
Manfield


CAT
DOG

KEY ML TERMINOLOGIES

What is (supervised) machine learning? Concisely put, it is the following:

- ML systems learn how to combine input to produce useful predictions on never-before-seen data.

Let's explore fundamental machine learning terminology.

Labels

A **label** is the thing we're predicting—the `y` variable in simple linear regression. The label could be the future price of wheat, the kind of animal shown in a picture, the meaning of an audio clip, or just about anything.

A good part or a bad part

KEY ML TERMINOLOGIES

- FEATURES

A **feature** is an input variable—the `x` variable in simple linear regression. A simple machine learning project might use a single feature, while a more sophisticated machine learning project could use millions of features, specified as:

$$x_1, x_2, \ldots x_N$$

In the spam detector example, the features could include the following:

- words in the email text
- sender's address
- time of day the email was sent
- email contains the phrase "one weird trick."

	spelling	special	unknown	emojis	Repeat	freq	Pri	Label
1	1	1	1	1	0	81	5	1
2	0	0	1	1	0			1
3								
4	0	0	0	0	1			0

Lets make synthetic data for BINARY CLASSIFICATION DATA for Cat & Dog seen in earlier slides

KEY ML TERMINOLOGIES

Examples

An **example** is a particular instance of data, **x**. (We put **x** in boldface to indicate that it is a vector.) We break examples into two categories:

- labeled examples
- unlabeled examples

A **labeled example** includes both feature(s) and the label. That is:

```
labeled examples: {features, label}: (x, y)
```

Make an example from mfr perspective

Use labeled examples to **train** the model. In our spam detector example, the labeled examples would be individual emails that users have explicitly marked as "spam" or "not spam."

We just made a data set showing labled examples for cat & dog

DATA

- Material type,
- Ductility,
- Hardness,
- Impact resistance,
- Fracture toughness,
- Tensile strength,
- Shear strength,
- Metal thickness,
- No of products

→ Time to Fail

We can create a more intrinsic dataset requiring more effort ☺

KEY ML TERMINOLOGIES

An **unlabeled example** contains features but not the label. That is:

```
unlabeled examples: {features, ?}: (x, ?)
```

What will be unlabeled example for cat/dog classification problem?

Once we've trained our model with labeled examples, we use that model to predict the label on unlabeled examples. In the spam detector, unlabeled examples are new emails that humans haven't yet labeled.

KEY ML TERMINOLOGIES

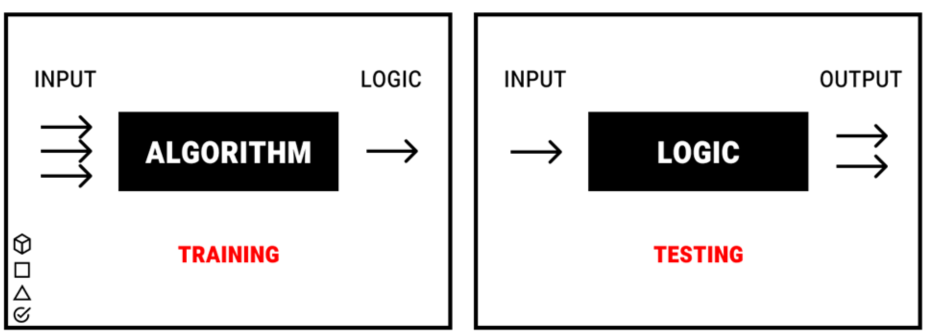
Models

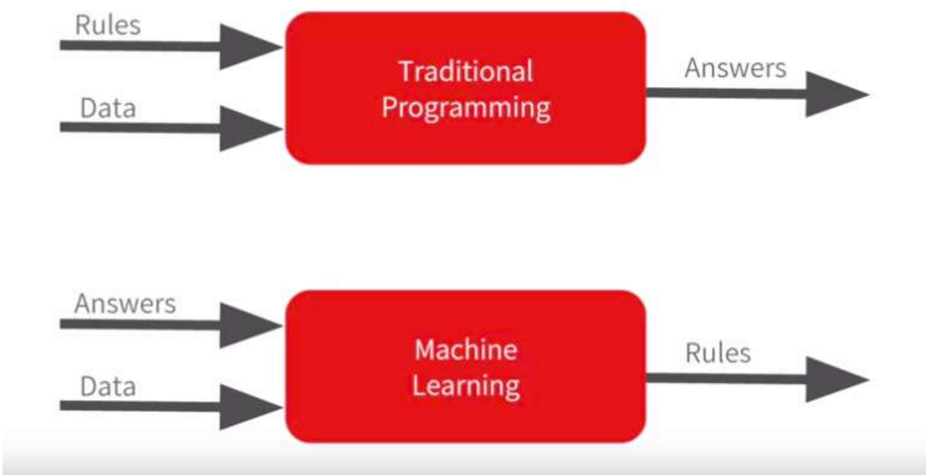
A model defines the relationship between features and label. For example, a spam detection model might associate certain features strongly with "spam". Let's highlight two phases of a model's life:

- **Training** means creating or **learning** the model. That is, you show the model labeled examples and enable the model to gradually learn the relationships between features and label.
- **Inference** means applying the trained model to unlabeled examples. That is, you use the trained model to make useful predictions (`y'`). For example, during inference, you can predict

Time of Failure
Cat or dog

for new unlabeled examples.





KEY ML TERMINOLOGIES

Regression vs. classification

A **regression** model predicts continuous values. For example, regression models make predictions that answer questions like the following:

- What is the value of a house in California?
- What is the probability that a user will click on this ad?

A **classification** model predicts discrete values. For example, classification models make predictions that answer questions like the following:

- Is a given email message spam or not spam?
- Is this an image of a dog, a cat, or a hamster?

What is type of problem that we took on till now?
Classification more common, lets do a regression task

LINEAR REGRESSION

- Crickets (an insect species) chirp more frequently on hotter days than on cooler days.
- Using this data, you want to explore this relationship

LINEAR REGRESSION

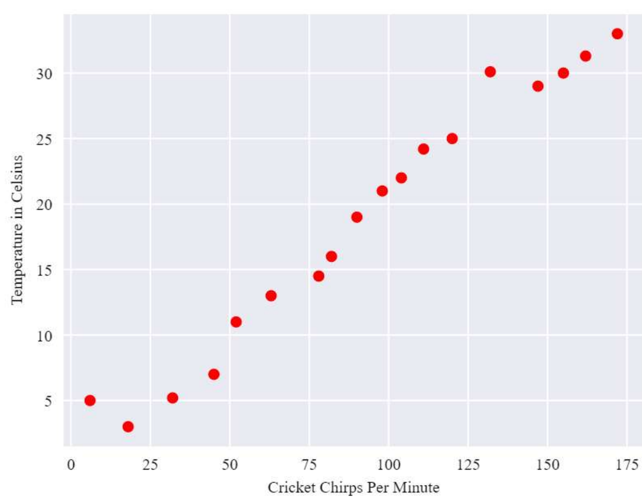


Figure 1. Chirps per Minute vs. Temperature in Celsius.

LINEAR REGRESSION

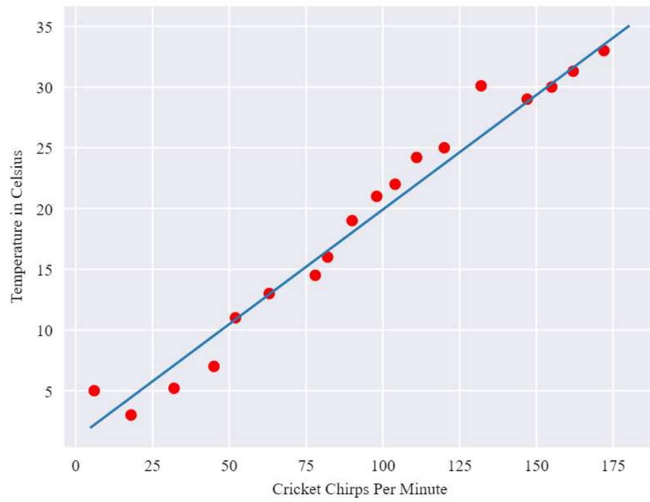


Figure 2. A linear relationship.

LINEAR REGRESSION

True, the line doesn't pass through every dot, but the line does clearly show the relationship between chirps and temperature. Using the equation for a line, you could write down this relationship as follows:

$$y = mx + b$$

where:

- y is the temperature in Celsius—the value we're trying to predict.
- m is the slope of the line.
- x is the number of chirps per minute—the value of our input feature.
- b is the y-intercept.

LINEAR REGRESSION

By convention in machine learning, you'll write the equation for a model slightly differently:

$$y' = b + w_1 x_1$$

where:

- y' is the predicted label (a desired output).
- b is the bias (the y-intercept), sometimes referred to as w_0.
- w_1 is the weight of feature 1. Weight is the same concept as the "slope" m in the traditional equation of a line.
- x_1 is a feature (a known input).

To **infer** (predict) the temperature y' for a new chirps-per-minute value x_1, just substitute the x_1 value into this model.

Although this model uses only one feature, a more sophisticated model might rely on multiple features, each having a separate weight (w_1, w_2, etc.). For example, a model that relies on three features might look as follows:

$$y' = b + w_1 x_1 + w_2 x_2 + w_3 x_3$$

Like human

DATA, AI ALGO, DATA

Databases
Kaggle
UCI
Github
Private sources

Data → AI Algorithm → Data

REMIND TO SHOW
AFTER GC exercise