Different types of Learning

Machine Learning Unit 3

Sudeshna Sarkar

Broad types of machine learning

- Supervised Learning
  - Training Data with labels: X,y (pre-classified)
  - Given an observation x, what is the best label for y?
- Unsupervised learning
  - Training Data without labels: X
  - Given a set of x's, find hidden structure
- Semi-supervised Learning
  - Training Data + some Labels
- Reinforcement Learning
  - Given: observations and periodic rewards as the agent takes sequential action in an environment
  - Determine optimum policy

Supervised Learning

- Given data containing the inputs and outputs:

Training Data:

$\{(\overline{x_1}, y_1), (\overline{x_2}, y_2), \ldots, (\overline{x_m}, y_m)\}$

- Learn a function $f(x)$ to predict y given x

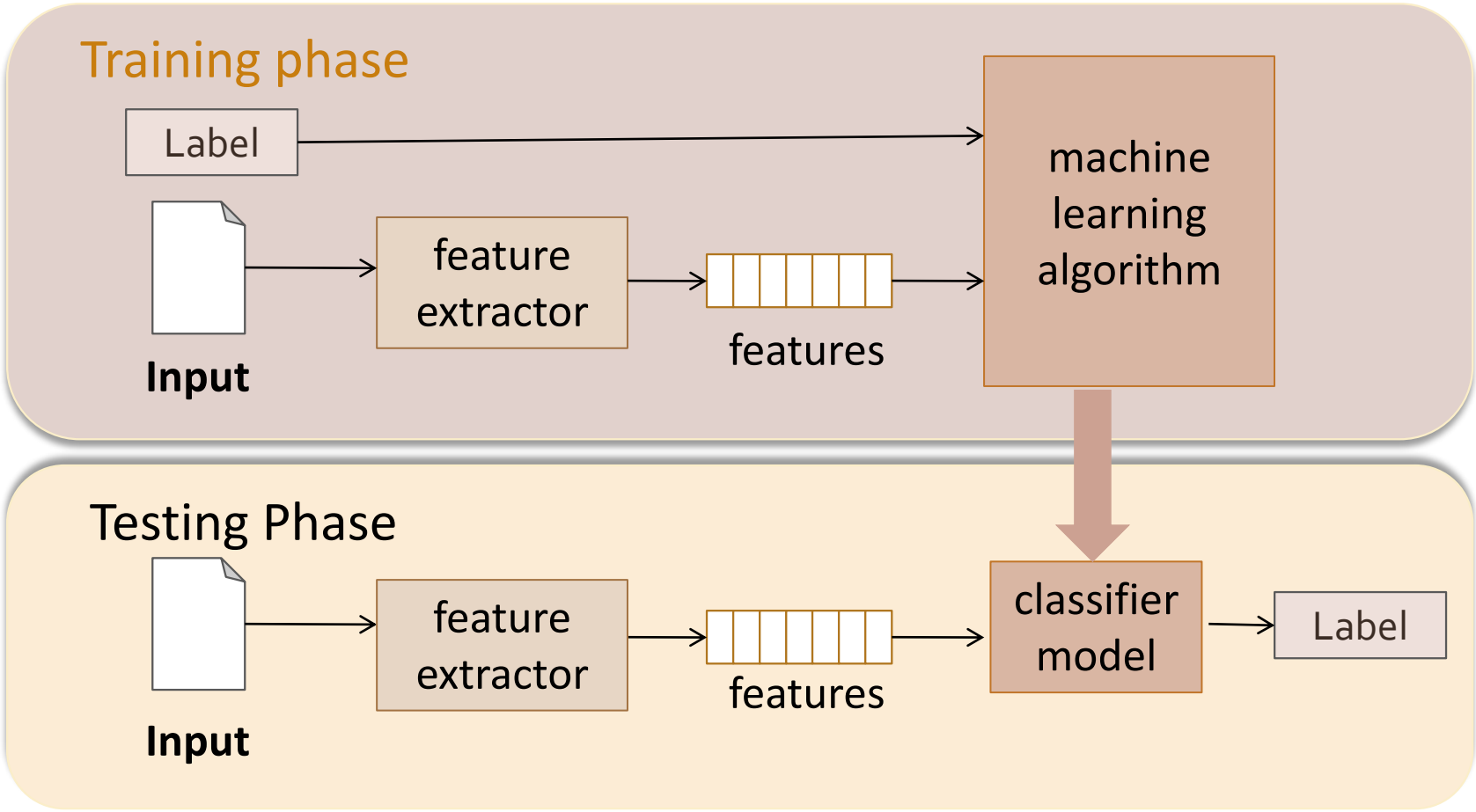
$\bar{X}$	Y
$\overline{x_1}$	y_1
$\overline{x_2}$	y_2
…	..
$\overline{x_m}$	y_m

Model → $f_\theta(\bar{x}) \rightarrow y$

Parameters

Training: Learn the model from the Training Data

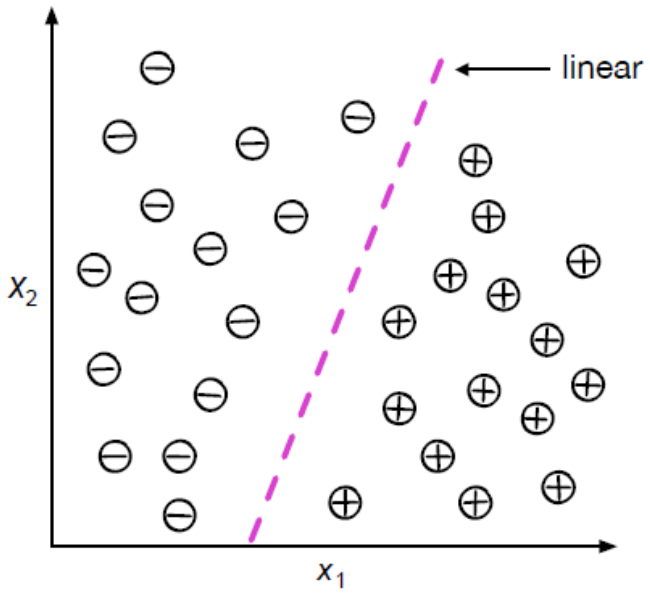
Given Test instance $\overline{x'}$, predict $y' = f_\theta(\overline{x'})$


Training phase
Label
machine learning algorithm
feature extractor
features
Input
Testing Phase
feature extractor
features
classifier model
Label
Input

Supervised Learning

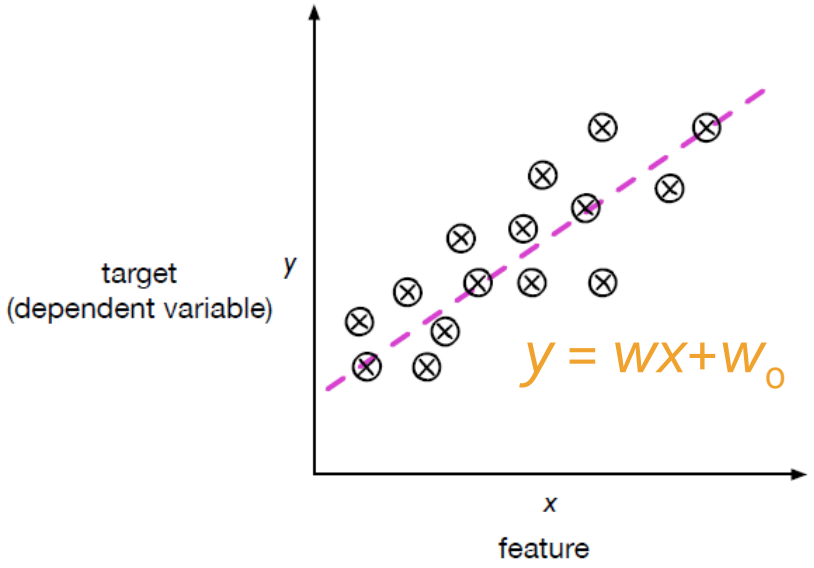
Classification

Y is categorical/ discrete



Regression

Y is numeric / continuous



Example Tasks

Classification

- Object identification from images
- Email classification
- Whether stock price will be up or down next month

Regression

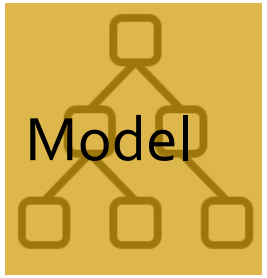
- House price prediction
- Stock price prediction
- Precipitation forecasting
- What is the probability that the user will buy this phone?

Supervised Learning
Classification Example

Training Samples

Color	Sleeve	Pattern	Collar	Like
Red	Half	Printed	No	Y
White	Half	Solid	Yes	N
…				..
White	Full	Solid	Yes	Y

Train a model to minimize loss



Test Instances

Color	Sleeve	Pattern	Collar	Like
Blue	Full	Printed	No	?
White	Half	Solid	No	?
Red	Half	Stripe	Yes	?
Red	Full	Solid	Yes	?

Probabilistic Classification

Color	Sleeve	Pattern	Collar	Category
Red	Half	Printed	No	Casual
White	Half	Solid	Yes	Lounge
…				..
White	Full	Solid	Yes	Formal

Predict a probability distribution over the set of classes **Pr (Y|X)**

Color	Sleeve	Pattern	Collar	Formal	Casual	Party
Blue	Full	Printed	No	0.4	0.15	0.45
White	Half	Solid	No	0.2	0.7	0.1
Red	Half	Stripe	Yes	0.1	0.2	0.7
Red	Full	Solid	Yes	0.5	0.1	0.4

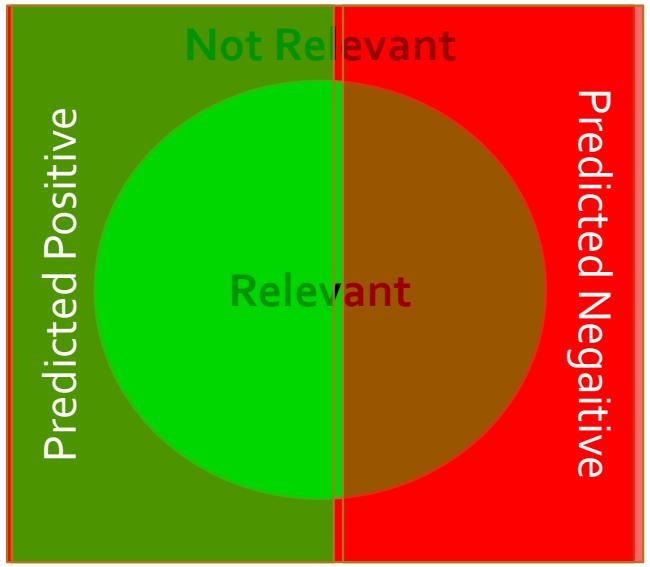
Evaluation for Classification problems

- Accuracy $= \frac{\text{\# correctly classified}}{\text{\# all test examples}}$

$= \frac{\#\text{predicted true pos} + \#\text{predicted true } neg}{\#\text{all test examples}}$

Precision $= \frac{\text{\# predicted true pos}}{\text{\# predicted pos}}$

Recall $= \frac{\text{\# predicted true pos}}{\text{\# True pos}}$



Loss Function
Classification problems

Loss indicates how bad the model's prediction is.

1. Fraction of Misclassifications

$$Error = \sum_{i=1}^{m} \frac{I(y_i \neq \hat{y}_i)}{m}$$

2. Logarithmic Loss: Maximize the log likelihood. For a loss function, minimize the negative log likelihood of the correct class:

$$L_i = -\log(P(Y = y_i | X = x_i))$$

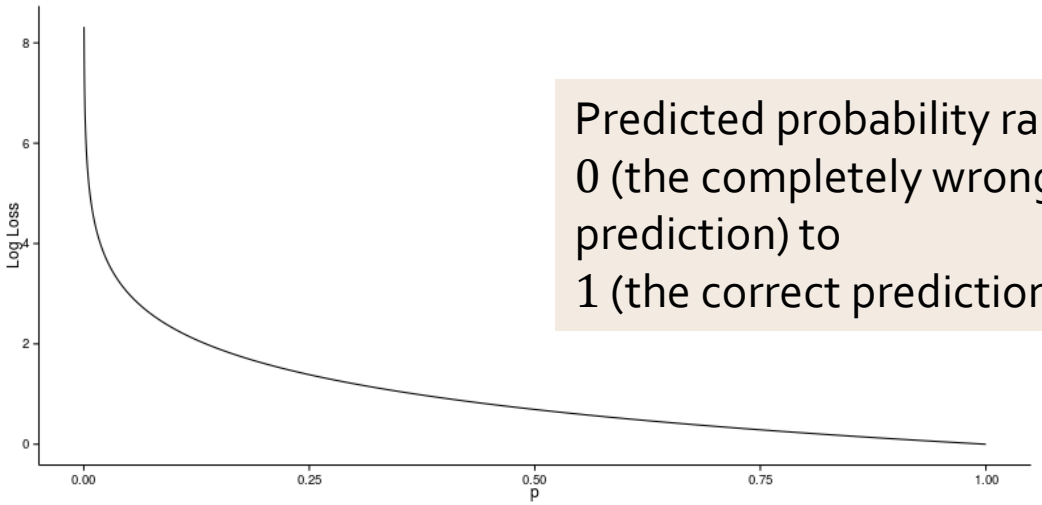
Logarithmic Loss Function

2. Logarithmic Loss:

$$L_i = -\log_M(P(Y = y_i|X = x_i))$$

$$L = \sum_{c=1} y_{oc} \log(p_{oc})$$

M - number of classes
y - binary indicator (0 or 1) if class label c is the correct classification for observation o
p - predicted probability observation o is of class c



Predicted probability ranges from 0 (the completely wrong prediction) to 1 (the correct prediction)

2. Evaluation for regression problem

- Mean Squared error

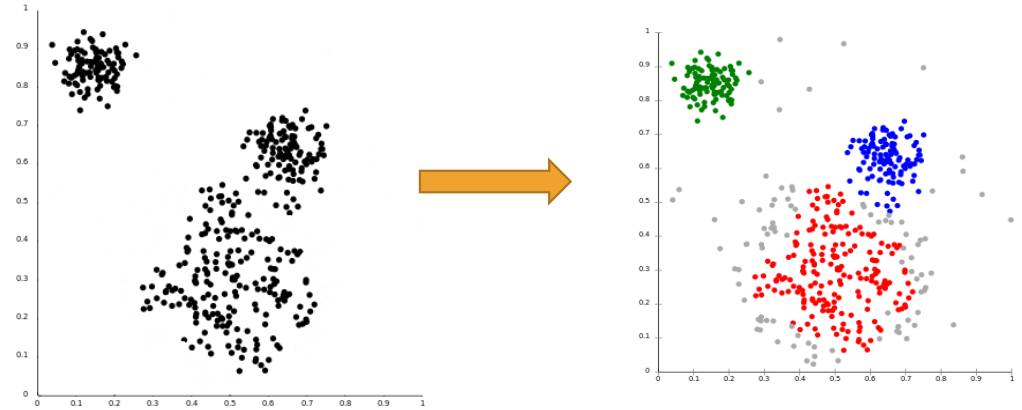
$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

Predicted value

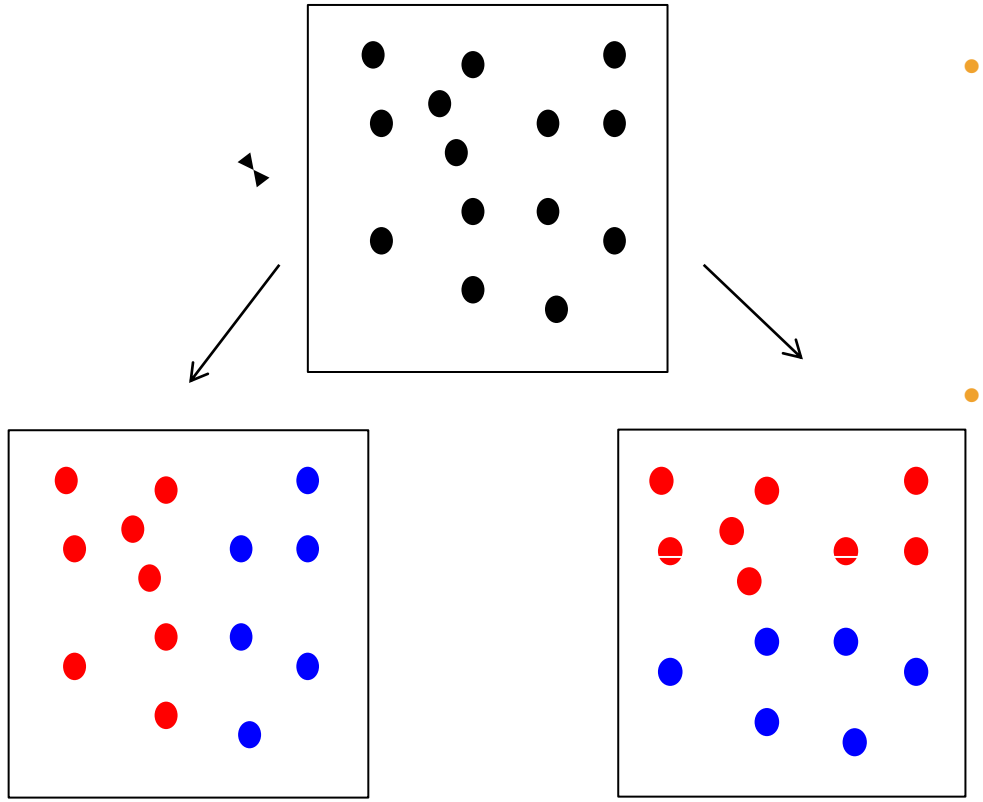
True value

Unsupervised Learning (Clustering)

- Given $\{\overline{x_1}, \overline{x_2}, \ldots \overline{x_m}, \}$ without labels
- Find hidden structure in the data
  - Clustering
  - Dimensionality Reduction
- Clustering: Grouping similar objects

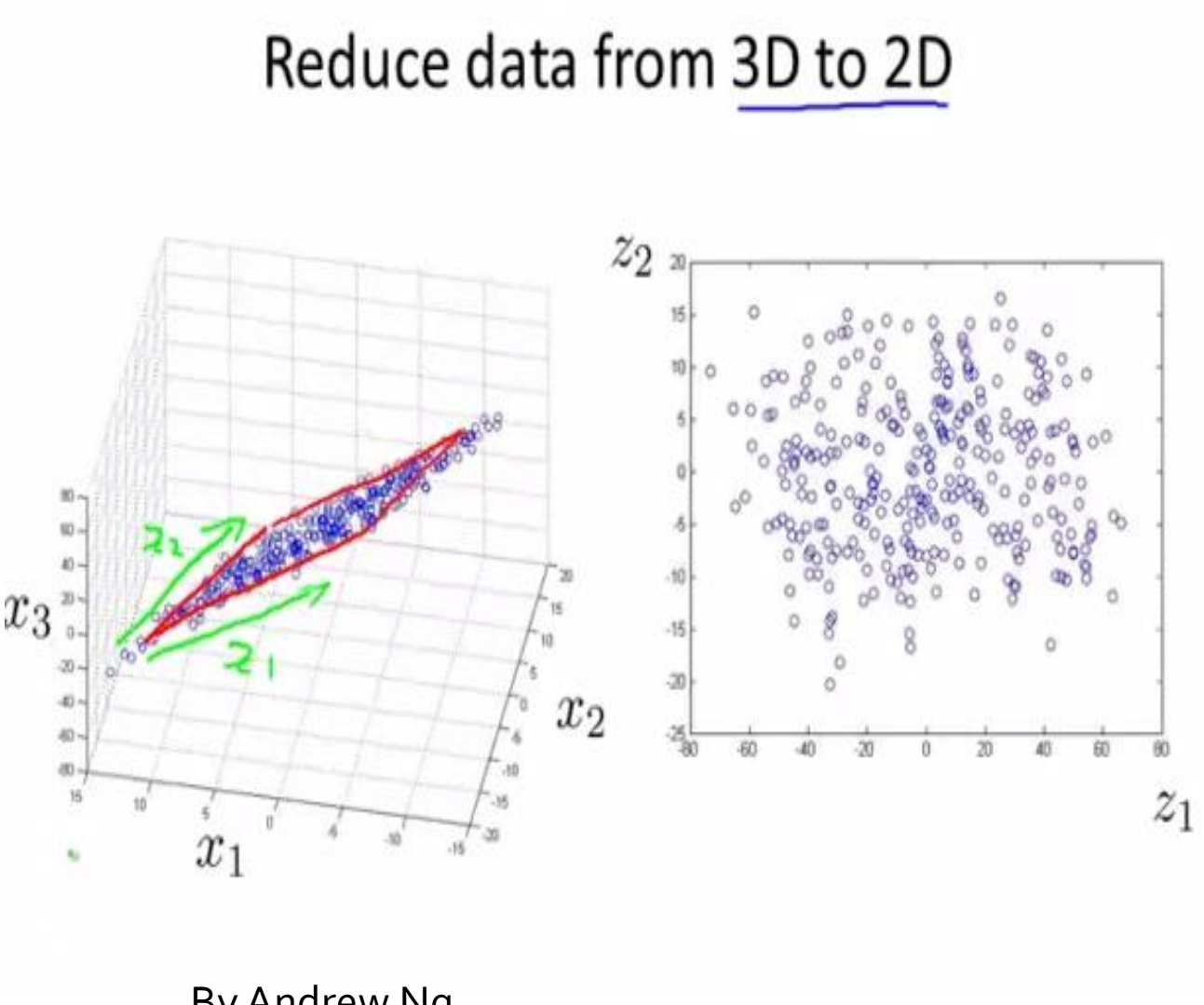


Clustering Problems



- How to evaluate clustering?
- Internal Evaluation:
  - Intra-cluster distances are minimized
  - Inter-cluster distances are maximized
- External Evaluation

Dimensionality Reduction



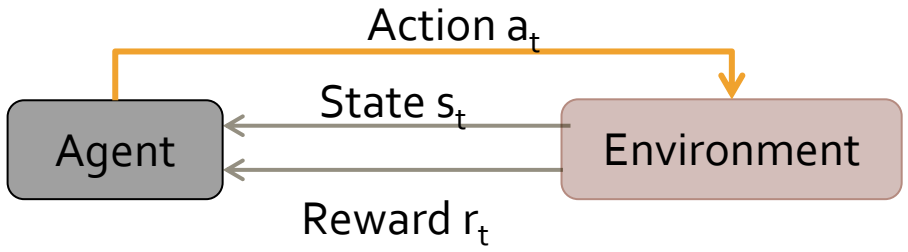
By Andrew Ng

Semi-Supervised Learning

- Supervised learning + Additional unlabeled data
- Unsupervised learning + Additional labeled data
- Learning Algorithm:
  - Start from the labeled data to build an initial classifier
  - Use the unlabeled data to enhance the model

Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy.



- Receive feedback in the form of rewards
- Agent's utility is defined by the reward function
- Must (learn to) act so as to maximize expected rewards

- Examples:
  - Game playing (Go)
  - Robot grasping
  - Controlling aircraft and robotic motion
  - Dynamic Pricing
  - Personalized recommendation

Goal: Constantly learn to make 'optimal' predictions based on real-time feedback from past predictions

Why Machine Learning?

- Human expertise not sacrosanct
  - Pricing, Promotions
- Human expertise cannot be coded
  - face/handwriting/speech recognition
  - driving a car, flying a plane
- Rapidly changing situation
  - Credit scoring, financial modeling
  - Fraud detection
- Need for personalization
  - Product recommendation
- Too much Data