CS 480

Introduction to Artificial Intelligence

November 4th, 2021

Announcements / Reminders

- **Programming Assignment #01:**
  - **due: ~~October 17th~~ ~~October 22th~~ ~~October 24th~~ ~~November 3rd~~,November 5th, 11:00 PM CST**
- **Programming Assignment #02:**
  - **TBA**
- **Written Assignment #03:**
  - **TBA**

CORRECTION: Expected Action Utility

The expected utility of an action a given the evidence is the average utility value of all possible outcomes s' of action a, weighted by their probability (belief) of occurence:

$$\mathrm{EU}(a) = \sum_{s'} \sum_{s} \mathrm{P}(s) * \mathrm{P}(s' \mid s, a) * \boxed{\mathrm{U}(s')} = \sum_{s'} \mathrm{P}(Result(a) = s') * \boxed{\mathrm{U}(s')}$$

Rational agent should choose an action that maximizes the expected utility:

$$\text{chosen action} = \underset{a}{\text{argmax}}\ \mathrm{EU}(a)$$

Syllabus: In Progress / Remaining

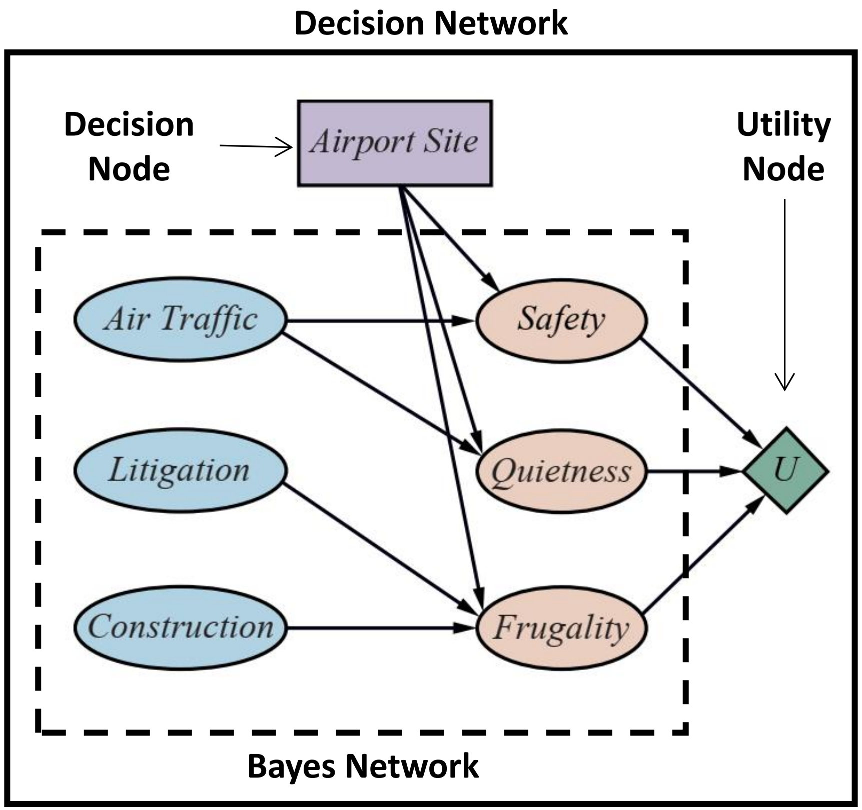
- Making Simple Decisions [Chapter 16]
- Making Complex Decisions [Chapter 17]
- Learning From Examples [Chapter 19]
- Deep Learning [Chapter 21]
- Reinforcement Learning [Chapter 22]
- Philosophy, Ethics, and Safety of AI [Chapter 27]
- The Future of AI [Chapter 28]

Plan for Today

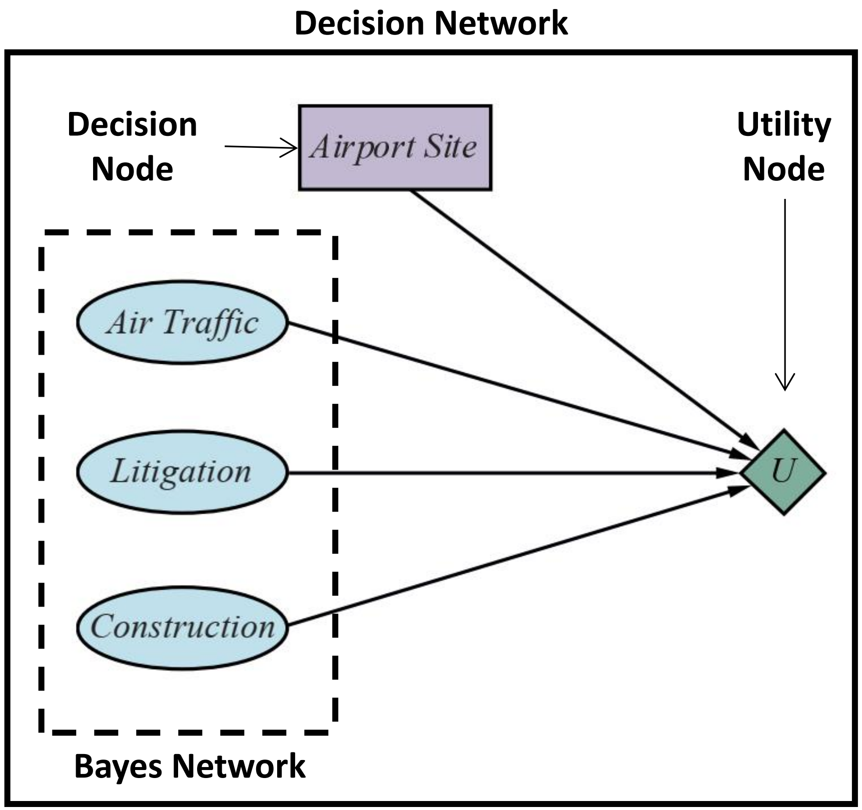
- **Making simple decisions**
- **Casual Introduction to Machine Learning**

(Single-Stage) Decision Networks

General Structure

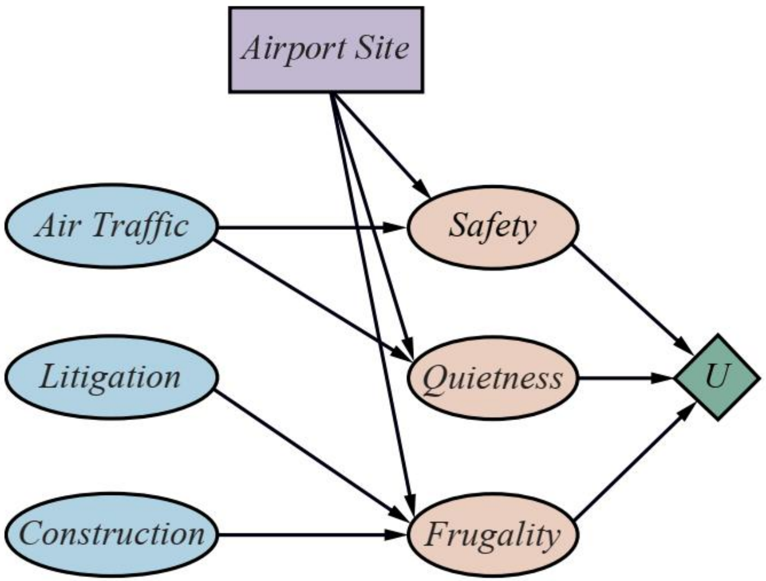


Simplified Structure



(Single-Stage) Decision Networks

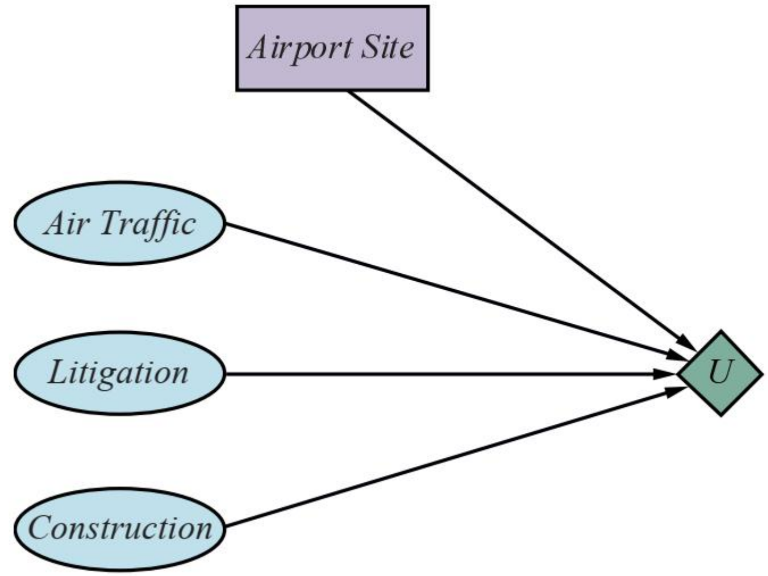
General Structure



Utility Table

S	low	low	low	low	high	high	high	high
Q	low	low	high	high	low	low	high	high
F	low	high	low	high	low	high	low	high
U	10	20	5	50	70	150	100	200

Simplified Structure



Action-Utility Table (not all columns shown)

AT	low	low	low	---	---	high	high	high
L	low	low	high	---	---	low	high	high
C	low	high	low	---	---	high	low	high
AS	A	A		---	---	B	B	B
U	10	20	5	---	---	150	100	200

Decision Network: Evaluation

The algorithm for decision network evaluation is as follows:

1. **Set the evidence variables for the current state**
2. **For each possible value** a **of decision node:**
   a. **Set the decision node to that value**
   b. **Calculate the posterior probabilities for the parent nodes of the utility node**
   c. **Calculate the utility for the action / value** a
3. **Return the action with highest utility**

Agent's Decisions

Recall that agent ACTIONS change the state:

- **if we are in state** s
- **action** a **is expected to**
- **lead to another state** s' **(outcome)**

Given uncertainty about the current state s **and action outcome** s' **we need to define the following:**

- **probability (belief) of being in state** s**:** $P(s)$
- **probability (belief) of action** a **leading to outcome** s'**:** $P(s' \mid s, a)$

Now:

$$P(s' \mid s, a) = P(\textsc{Result}(a) = s') = \sum_s P(s) * P(s' \mid s, a)$$

Expected Action Utility

The expected utility of an action a given the evidence is the average utility value of all possible outcomes s' of action a, weighted by their probability (belief) of occurence:

$$\mathrm{EU}(a) = \sum_{s'} \sum_{s} \mathrm{P}(s) * \mathrm{P}(s' \mid s, a) * \mathrm{U}(s') = \sum_{s'} \mathrm{P}(Result(a) = s') * \mathrm{U}(s')$$

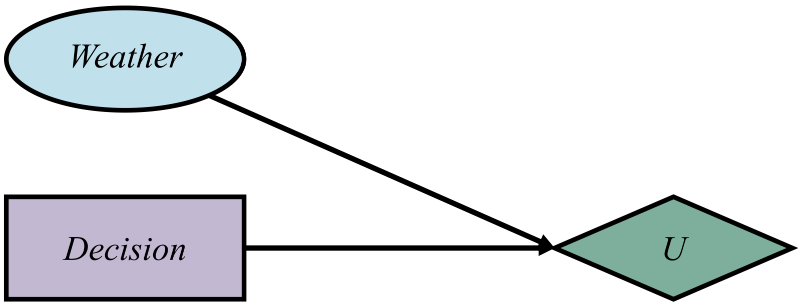
Rational agent should choose an action that maximizes the expected utility:

$$\text{chosen action} = \underset{a}{\operatorname{argmax}} \ \mathrm{EU}(a)$$

Decision Networks: Example

Decision: take umbrella

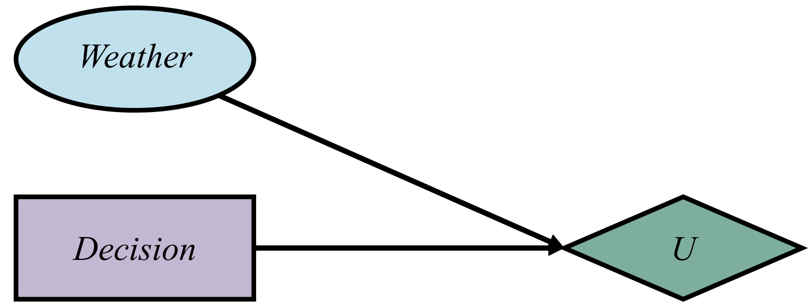
P(W=rain)	P(W=sun)
0.30	0.70



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Decision: leave umbrella

P(W=rain)	P(W=sun)
0.30	0.70



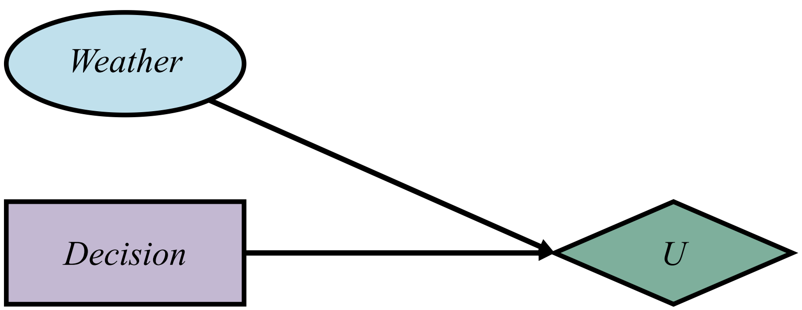
D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Decision Networks: Example

Decision: take umbrella

$$\mathrm{EU}(a) = \sum_{s'} \mathrm{P}(Result(a) = s') * \mathrm{U}(s')$$

P(W=rain)	P(W=sun)
0.30	0.70



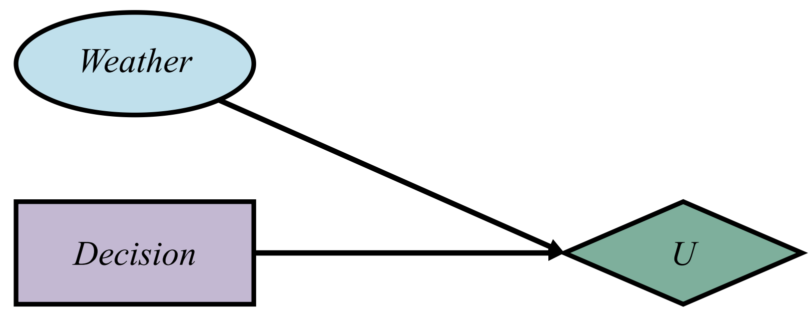
D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

$\mathrm{EU}(take) = ???$

Decision: leave umbrella

$$\mathrm{EU}(a) = \sum_{s'} \mathrm{P}(Result(a) = s') * \mathrm{U}(s')$$

P(W=rain)	P(W=sun)
0.30	0.70



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

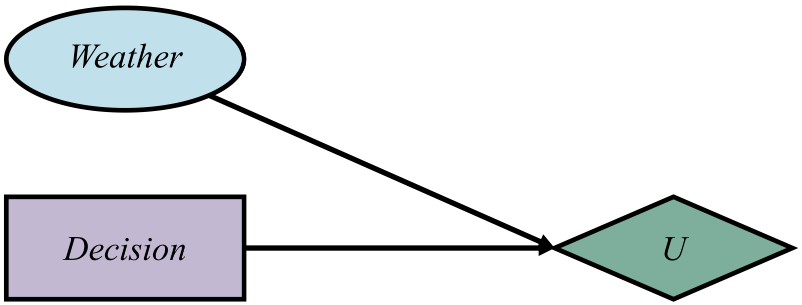
$\mathrm{EU}(leave) = ???$

Decision Networks: Example

Decision: take umbrella

$$EU(a) = \sum_{s'} P(Result(a) = s') * U(s')$$

P(W=rain)	P(W=sun)
0.30	0.70



S_1': D = take, W = sun
S_2': D = take, W = rain
EU(take) =
P(Result(take) = S_1')*U(S_1') +
P(Result(take) = S_2')*U(S_2') =
0.70 * 20 + 0.30 * 70 = 35

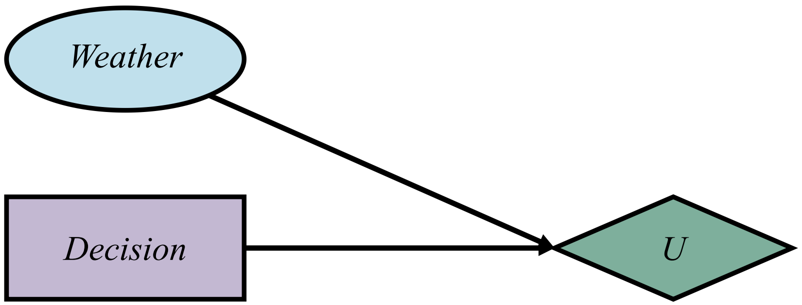
D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

$$EU(take) = 35$$

Decision: leave umbrella

$$EU(a) = \sum_{s'} P(Result(a) = s') * U(s')$$

P(W=rain)	P(W=sun)
0.30	0.70



S_3': D = leave, W = sun
S_4': D = leave, W = rain
EU(leave) =
P(Result(leave) = S_3')*U(S_3') +
P(Result(leave) = S_4')*U(S_4') =
0.70 * 100 + 0.30 * 0 = 70

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

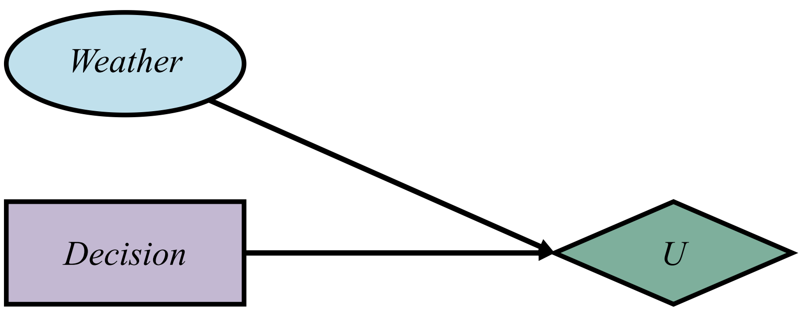
$$EU(leave) = 70$$

Decision Networks: Example

Which action to choose: take or leave Umbrella?

$$EU(a) = \sum_{s'} P(Result(a) = s') * U(s')$$

P(W=rain)	P(W=sun)
0.30	0.70

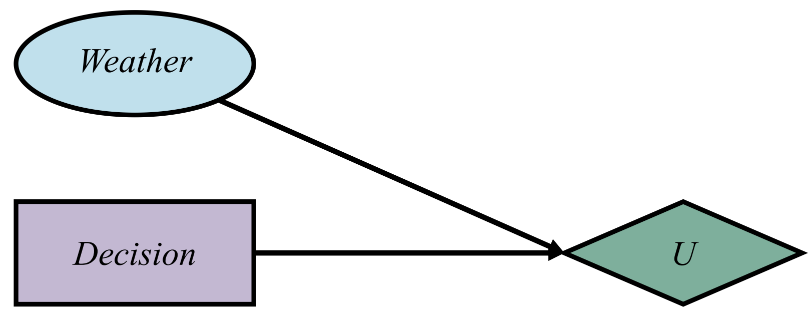


S_1': D = take, W = sun
S_2': D = take, W = rain
EU(take) =
P(Result(take) = S_1')*U(S_1') +
P(Result(take) = S_2')*U(S_2') =
0.70 * 20 + 0.30 * 70 = 35

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

$$EU(a) = \sum_{s'} P(Result(a) = s') * U(s')$$

P(W=rain)	P(W=sun)
0.30	0.70



S_3': D = leave, W = sun
S_4': D = leave, W = rain
EU(leave) =
P(Result(leave) = S_3')*U(S_3') +
P(Result(leave) = S_4') U(S_4') =
0.70 * 100 + 0.30 * 0 = 70

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

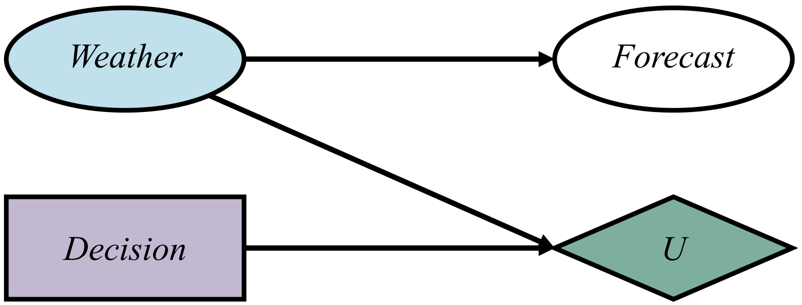
$$action = \underset{a}{\operatorname{argmax}} EU(a) \mid \max(EU(take), \underline{EU(leave)}) = \max(35, \mathbf{70}) \rightarrow leave$$

Decision Networks: Example

Decision: take umbrella

P(W=rain)	P(W=sun)
0.30	0.70

P(F=sun)	P(F=rain)
0.59	0.41

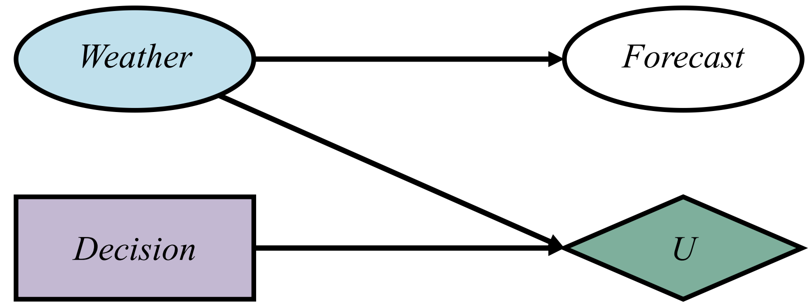


D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Decision: leave umbrella

P(W=rain)	P(W=sun)
0.30	0.70

P(F=sun)	P(F=rain)
0.59	0.41



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

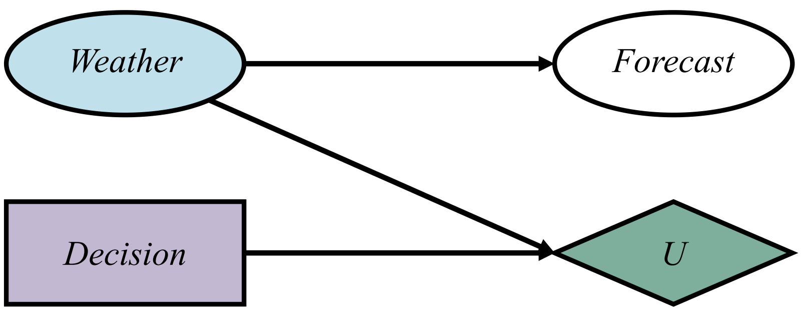
Decision Networks: Example

Decision: take umbrella given e

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(rain | F)	P(sun | F)
???	???

P(F=sun)	P(F=rain)
0.59	0.41



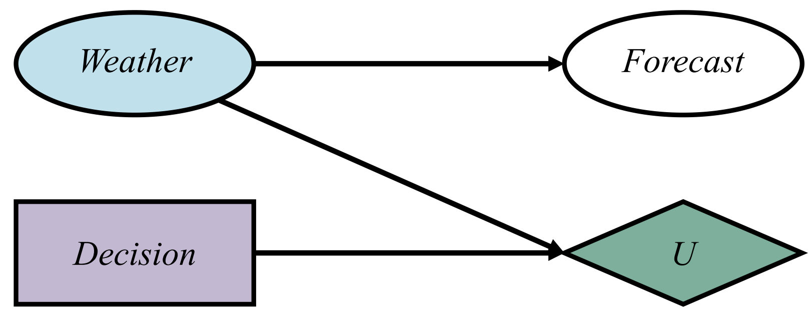
D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Decision: leave umbrella given e

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(rain | F)	P(sun | F)
???	???

P(F=sun)	P(F=rain)
0.59	0.41



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

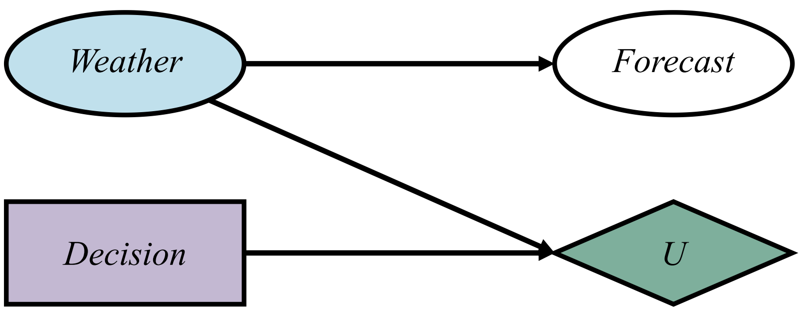
Decision Networks: Example

Decision: take umbrella given e

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(W=rain)	P(W=sun)
0.30	0.70

P(F=sun)	P(F=rain)
0.59	0.41



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

Conditional probabilities

Assume that we are given:

F	W	P(F|W)
sun	sun	0.80
rain	sun	0.20
sun	rain	0.10
rain	rain	0.90

By Bayes' Theorem:

$$P(W = sun \mid F = sun) = \frac{P(F = sun \mid W = sun) * P(W = sun)}{P(F = sun)} = \frac{0.80 * 0.70}{0.59} = 0.95$$

$$P(W = sun \mid F = rain) = \frac{P(F = rain \mid W = sun) * P(W = sun)}{P(F = rain)} = \frac{0.20 * 0.70}{0.41} = 0.34$$

$$P(W = rain \mid F = sun) = \frac{P(F = sun \mid W = rain) * P(W = rain)}{P(F = sun)} = \frac{0.10 * 0.30}{0.59} = 0.05$$

$$P(W = rain \mid F = rain) = \frac{P(F = rain \mid W = rain) * P(W = rain)}{P(F = rain)} = \frac{0.90 * 0.30}{0.41} = 0.66$$

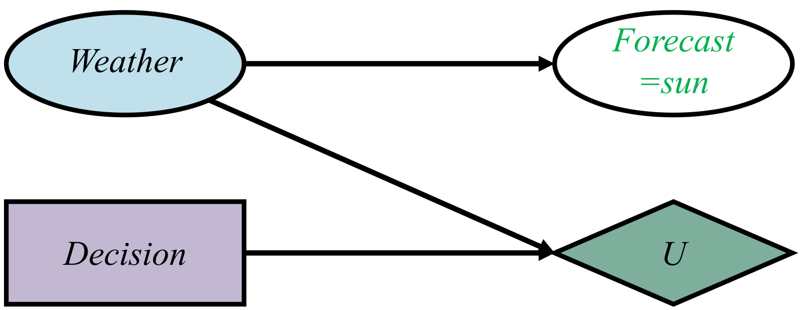
Decision Networks: Example

Decision: take umbrella given sun

$\mathrm{EU}(a \mid e) = \sum_{s'} \mathrm{P}(Result(a) = s' \mid e) * \mathrm{U}(s')$

P(rain | F)	P(sun | F)
0.05	0.95

P(F=sun)	P(F=rain)
0.59	0.41



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

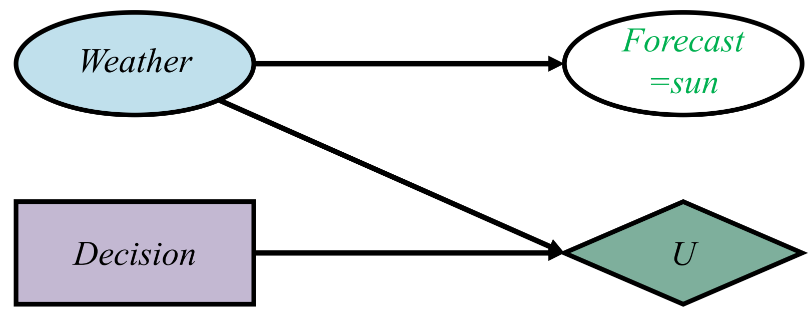
EU(take given sun forecast) = ? ? ?

Decision: leave umbrella given sun

$\mathrm{EU}(a \mid e) = \sum_{s'} \mathrm{P}(Result(a) = s' \mid e) * \mathrm{U}(s')$

P(rain | F)	P(sun | F)
0.05	0.95

P(F=sun)	P(F=rain)
0.59	0.41



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

EU(leave given sun forecast) = ? ? ?

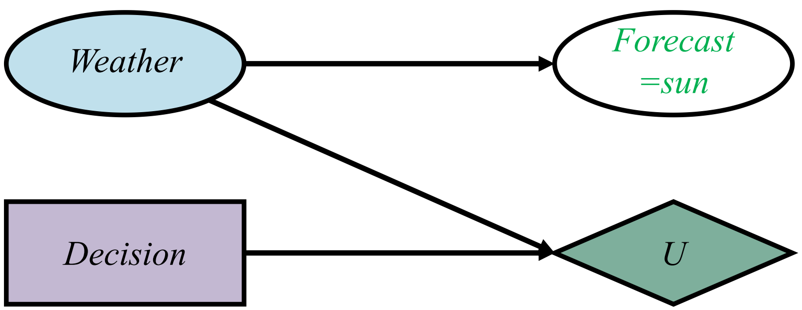
Decision Networks: Example

Decision: take umbrella given sun

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(rain | F)	P(sun | F)
0.05	0.95

P(F=sun)	P(F=rain)
0.59	0.41



S_1': D = take, W = sun
S_2': D = take, W = rain
EU(take) =
P(Result(take)=S_1'|e)*U(S_1') +
P(Result(take)=S_2'|e)*U(S_2') =
0.95 * 20 + 0.05 * 70 = 22.5

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

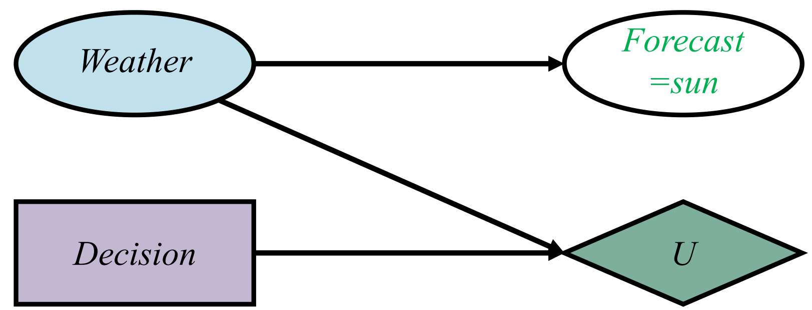
EU(take given sun forecast) = 22.5

Decision: leave umbrella given sun

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(rain | F)	P(sun | F)
0.05	0.95

P(F=sun)	P(F=rain)
0.59	0.41



S_3': D = leave, W = sun
S_4': D = leave, W = rain
EU(leave) =
P(Result(leave)=S_3'|e)*U(S_3') +
P(Result(leave)=S_4'|e)*U(S_4') =
0.95 * 100 + 0.05 * 0 = 95

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

EU(leave given sun forecast) = 95

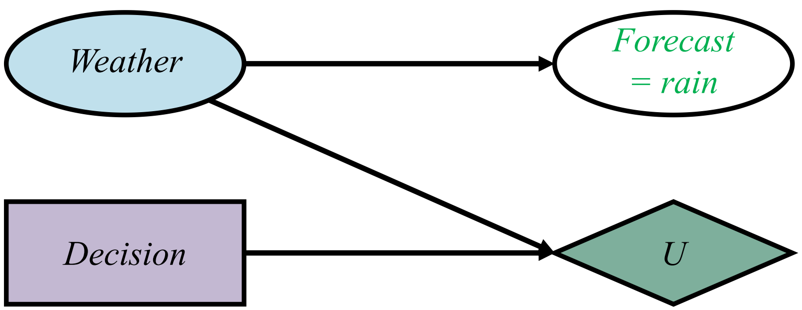
Decision Networks: Example

Decision: take umbrella given rain

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(rain | F)	P(sun | F)
0.66	0.34

P(F=sun)	P(F=rain)
0.59	**0.41**



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

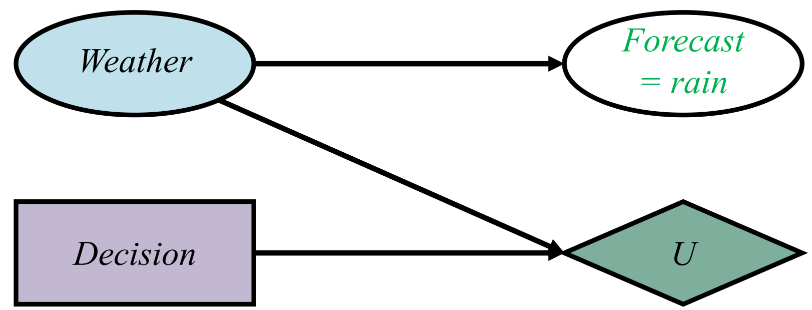
EU(take given rain forecast) = ? ? ?

Decision: leave umbrella given rain

$$EU(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$$

P(rain | F)	P(sun | F)
0.66	0.34

P(F=sun)	P(F=rain)
0.59	**0.41**



D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

EU(leave given rain forecast) = ? ? ?

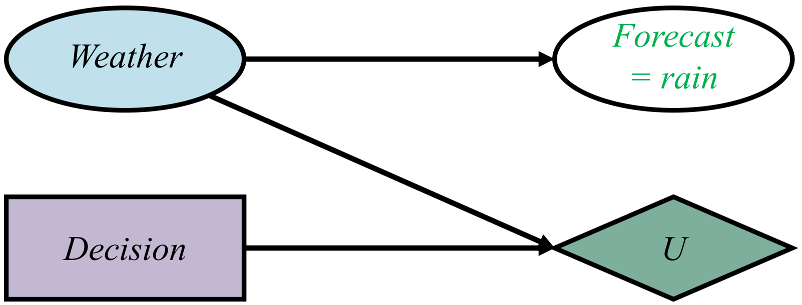
Decision Networks: Example

Decision: take umbrella given rain

$\text{EU}(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$

P(rain | F)	P(sun | F)
0.66	0.34

P(F=sun)	P(F=rain)
0.59	**0.41**



S_1': D = take, W = sun
S_2': D = take, W = rain
EU(take) =
P(Result(take)=S_1'|e)*U(S_1') +
P(Result(take)=S_2'|e)*U(S_2') =
0.34 * 20 + 0.66 * 70 = 53

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

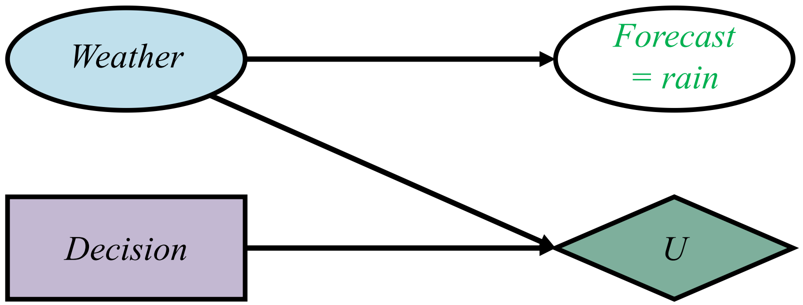
EU(take given rain forecast) = 53

Decision: leave umbrella given rain

$\text{EU}(a \mid e) = \sum_{s'} P(Result(a) = s' \mid e) * U(s')$

P(rain | F)	P(sun | F)
0.66	0.34

P(F=sun)	P(F=rain)
0.59	**0.41**



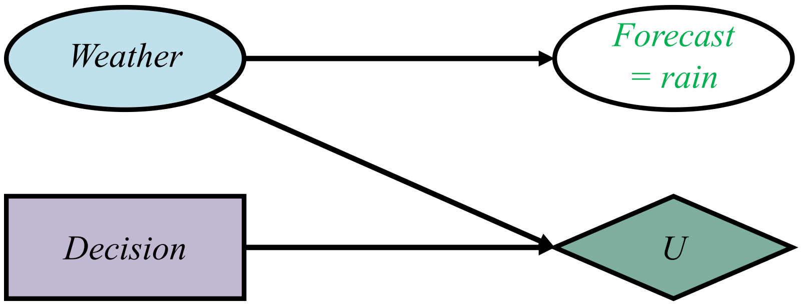
S_3': D = leave, W = sun
S_4': D = leave, W = rain
EU(leave) =
P(Result(leave)=S_3'|e)*U(S_3') +
P(Result(leave)=S_4'|e)*U(S_4') =
0.34 * 100 + 0.66 * 0 = 34

D	W	U
leave	sun	100
leave	rain	0
take	sun	20
take	rain	70

EU(leave given rain forecast) = 34

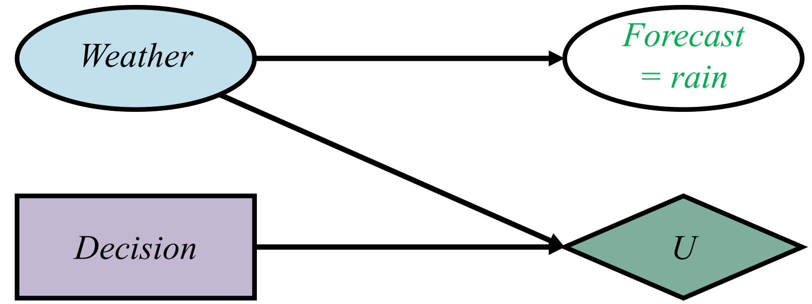
Decision Networks: Example

Decision: take umbrella given rain



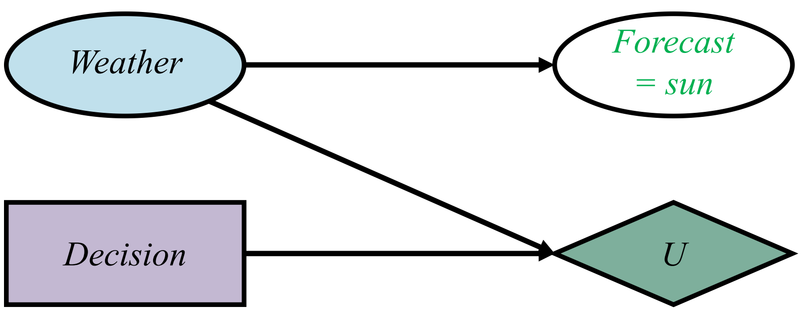
EU(take given rain forecast) = 53

Decision: leave umbrella given rain



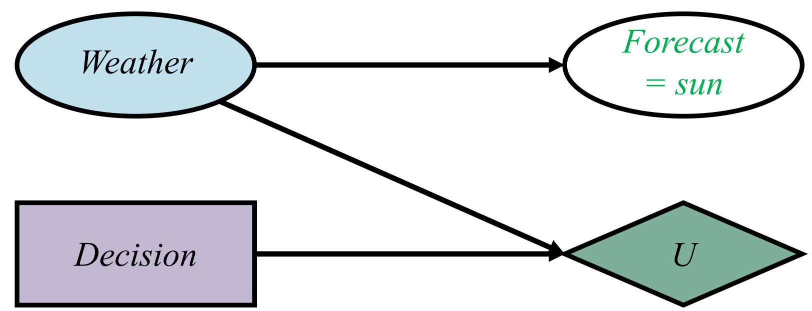
EU(leave given rain forecast) = 34

Decision: take umbrella given sun



EU(take given sun forecast) = 22.5

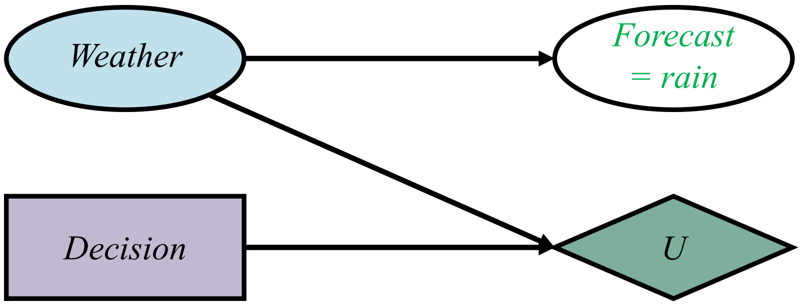
Decision: leave umbrella given sun



EU(leave given sun forecast) = 95

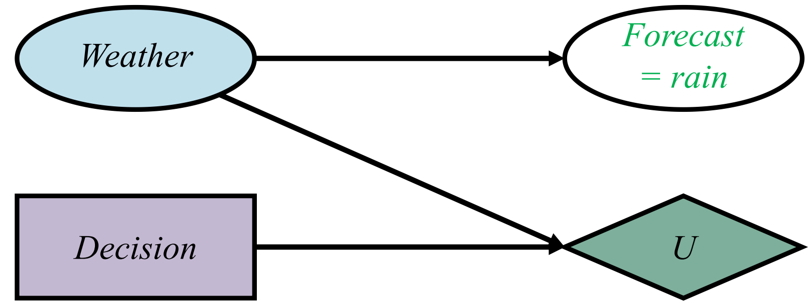
Decision Networks: Example

Decision:take umbrella given rain



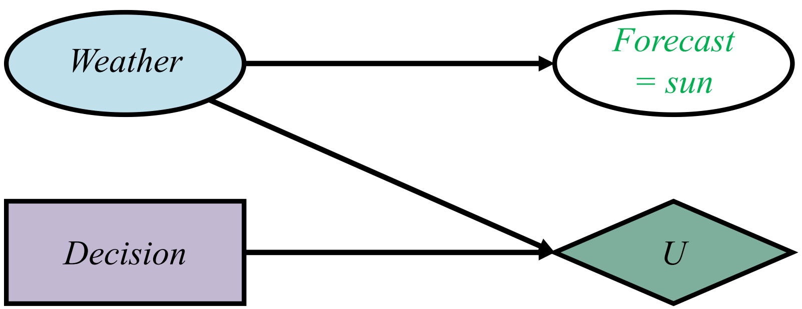
EU(take given rain forecast) = 53

Decision:leave umbrella given rain



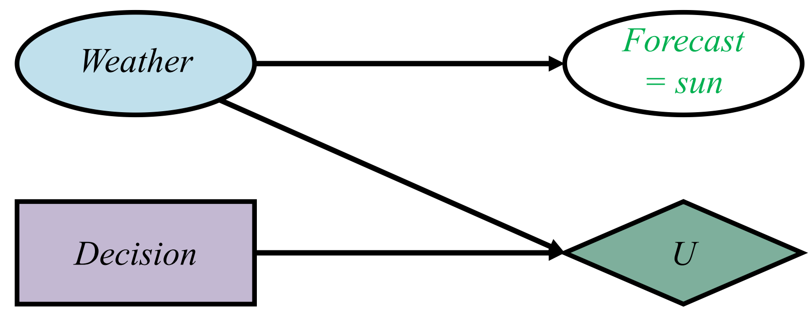
EU(leave given rain forecast) = 34

Decision:take umbrella given sun



EU(take given sun forecast) = 22.5

Decision:leave umbrella given sun



EU(leave given sun forecast) = 95

Value of Perfect Information

The value/utility of best action α without additional evidence (information) is :

$$\mathrm{MEU}(\alpha) = \max_{a} \sum_{s'} \mathrm{P}(Result(a) = s') * \mathrm{U}(s')$$

If we include new evidence/information ($\mathrm{E_j} = \mathrm{e_j}$) given by some variable Ej, value/utility of best action α becomes:

$$\mathrm{MEU}(a_{e_j} \mid e_j) = \max_{a} \sum_{s'} \mathrm{P}(Result(a) = s' \mid e_j) * \mathrm{U}(s')$$

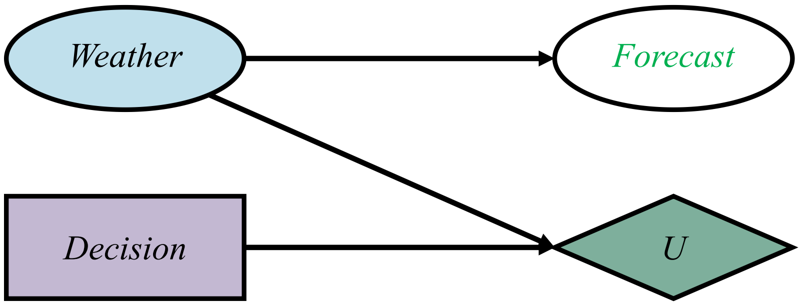
The value of additional evidence/information from Ej is:

$$\mathrm{VPI}(E_j) = \left(\sum_{e_j} \mathrm{P}(E_j = e_j) * \mathrm{MEU}(a_{e_j} \mid E_j = e_j)\right) - MEU(a)$$

using our current beliefs about the world.

Decision Network: Example

Decision network



The value of best action α without additional evidence

$$\mathrm{MEU}(\alpha) = \mathrm{MEU}(\mathrm{leave}) = 70$$

With evidence information ($E_j = e_j$) given by Forecast:

$$\mathrm{MEU}(a_{e_1} \mid e_1) = \mathrm{MEU}(\mathrm{take} \mid F = rain) = 53$$

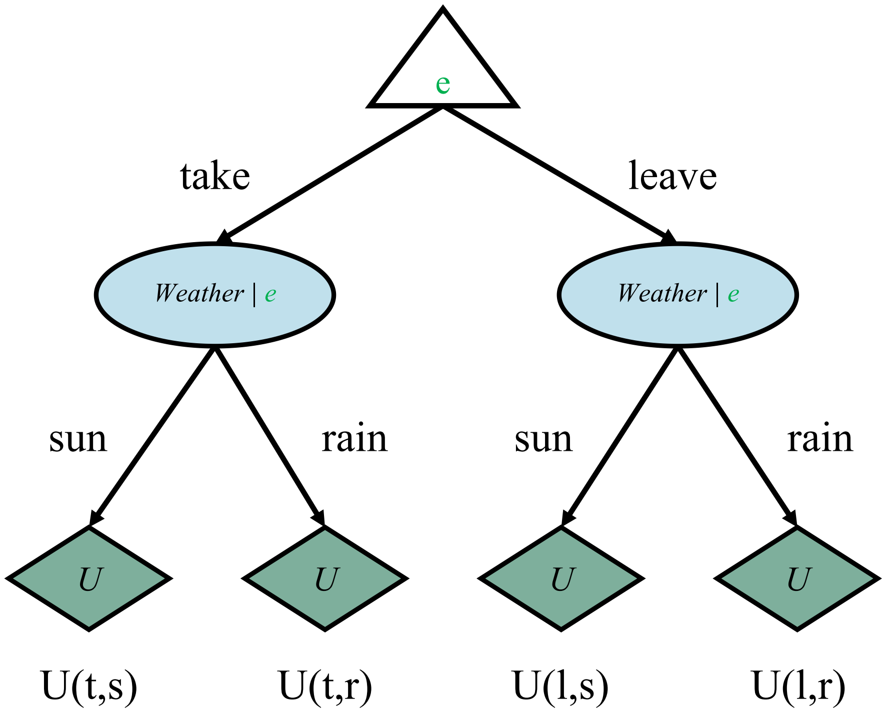
$$\mathrm{MEU}(a_{e_2} \mid e_2) = \mathrm{MEU}(\mathrm{leave} \mid F = sun) = 95$$

The value of additional evidence / information from F is:

$$\mathrm{VPI}(E_j) = \left(\sum_{e_j} \mathrm{P}(E_j = e_j) * \mathrm{MEU}(a_{e_j} \mid E_j = e_j)\right) - MEU(a)$$

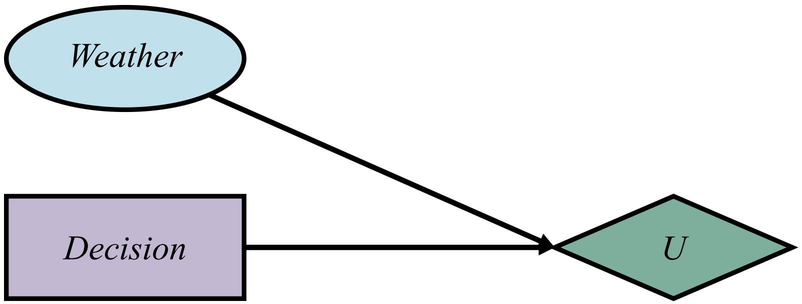
$$\mathrm{VPI}(F) = (\mathrm{P}(F = rain) * \mathrm{MEU}(take \mid F = rain) + \mathrm{P}(F = sun) * \mathrm{MEU}(\mathrm{leave} \mid F = sun)) - MEU(leave) = (0.41 * 53 + 0.59 * 95) - 70 = 7.78$$

Outcome tree



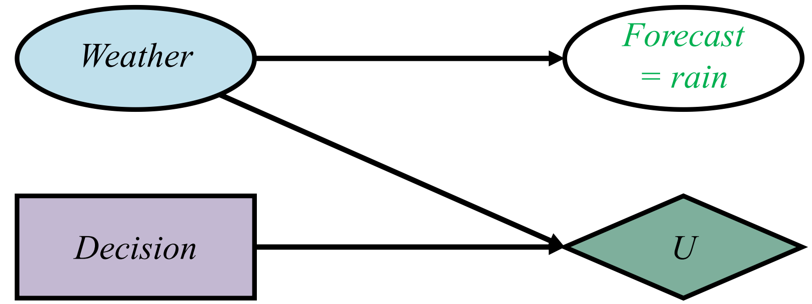
Decision Networks: Example

Decision: leave umbrella



$$\mathrm{EU}(\mathrm{leave}) = 70$$

Decision: take umbrella given rain



$$\mathrm{EU}(\mathrm{take\ given\ rain\ forecast}) = 53$$

The value of best action α without additional evidence

$$\mathrm{MEU}(\alpha) = \mathrm{MEU}(\mathrm{leave}) = 70$$

With evidence information ($\mathrm{E_j} = \mathrm{e_j}$) given by Forecast:

$$\mathrm{MEU}(a_{e_1} \mid e_1) = \mathrm{MEU}(\mathrm{take} \mid F = rain) = 53$$

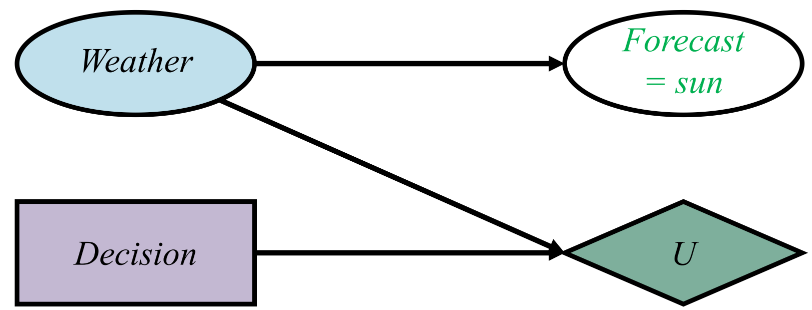
$$\mathrm{MEU}(a_{e_2} \mid e_2) = \mathrm{MEU}(\mathrm{leave} \mid F = sun) = 95$$

The value of additional evidence / information from F is:

$$\mathrm{VPI}(E_j) = \left(\sum_{e_j} \mathrm{P}(E_j = e_j) * \mathrm{MEU}(a_{e_j} \mid E_j = e_j)\right) - MEU(a)$$

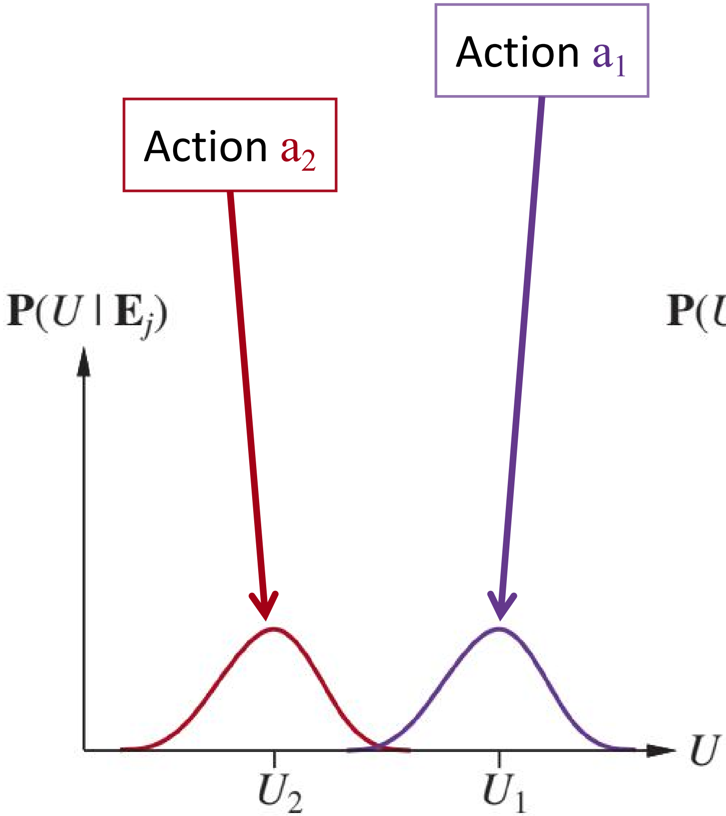
$$\mathrm{VPI}(F) = (\mathrm{P}(F = rain) * \mathrm{MEU}(take \mid F = rain) + \mathrm{P}(F = sun) * \mathrm{MEU}(\mathrm{leave} \mid F = sun)) - MEU(leave) = (0.41 * 53 + 0.59 * 95) - 70 = 7.78$$

Decision: leave umbrella given sun

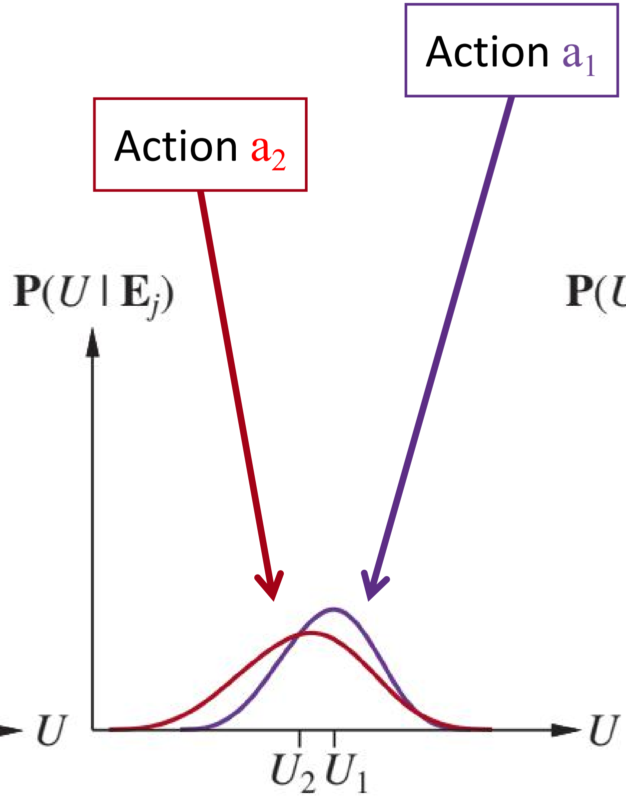


$$\mathrm{EU}(\mathrm{leave\ given\ sun\ forecast}) = 95$$

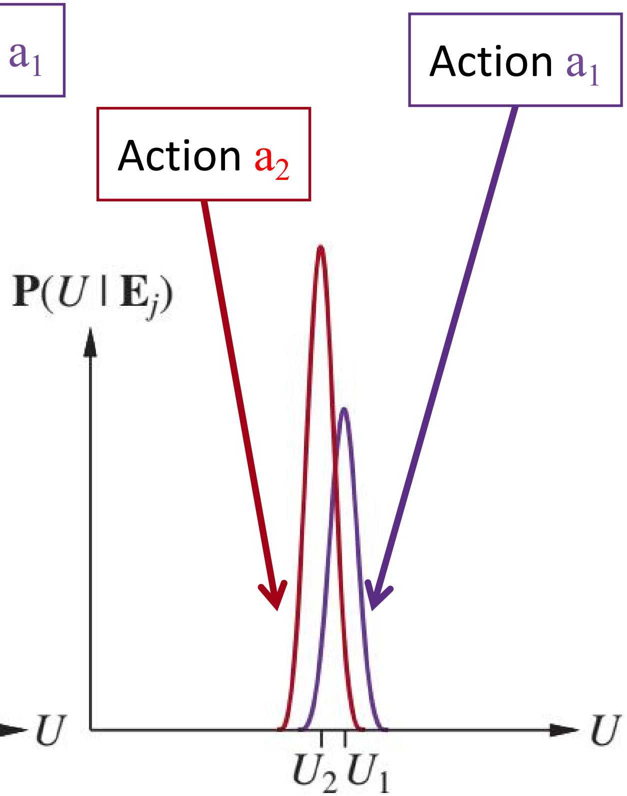
Utility & Value of Perfect Information



New information will not help here.



New information may help a lot here.



New information may help a bit here.

VPI Properties

Given a decision network with possible observations Ej (sources of new information / evidence):

- **The expected value of information is nonnegative:**

$$\forall_j \, \mathrm{VPI}(E_j) \geq 0$$

- **VPI is not additive:**

$$\mathrm{VPI}(E_j, E_k) \neq \mathrm{VPI}(E_j) + \mathrm{VPI}(E_k)$$

- **VPI is order-independent:**

$$\mathrm{VPI}(E_j,\ E_k) = VPI(E_j) \ + VPI(\ E_k \mid E_j) \ = VPI(E_k) \ + VPI(\ E_j \mid E_k) = VPI(E_k\ ,\ E_j)$$

Information Gathering Agent

function INFORMATION-GATHERING-AGENT(*percept*) **returns** an *action*
  persistent: D, a decision network

  integrate *percept* into D
  $j \leftarrow$ the value that maximizes $VPI(E_j)\ /\ C(E_j)$
  if $VPI(E_j)\ >\ C(E_j)$
    then return $Request(E_j)$
  else return the best action from D

How do you choose which media to share with others?

How do / did you learn things?

Machine Learning (ML)

Oxford English Dictionary Definition

Machine Learning:

A type of **artificial intelligence** in which computers **use huge amounts of data to learn how to do tasks** rather than being programmed to do them.

Source: https://www.oxfordlearnersdictionaries.com/us/definition/english/machine-learning

Traditional Programming vs ML

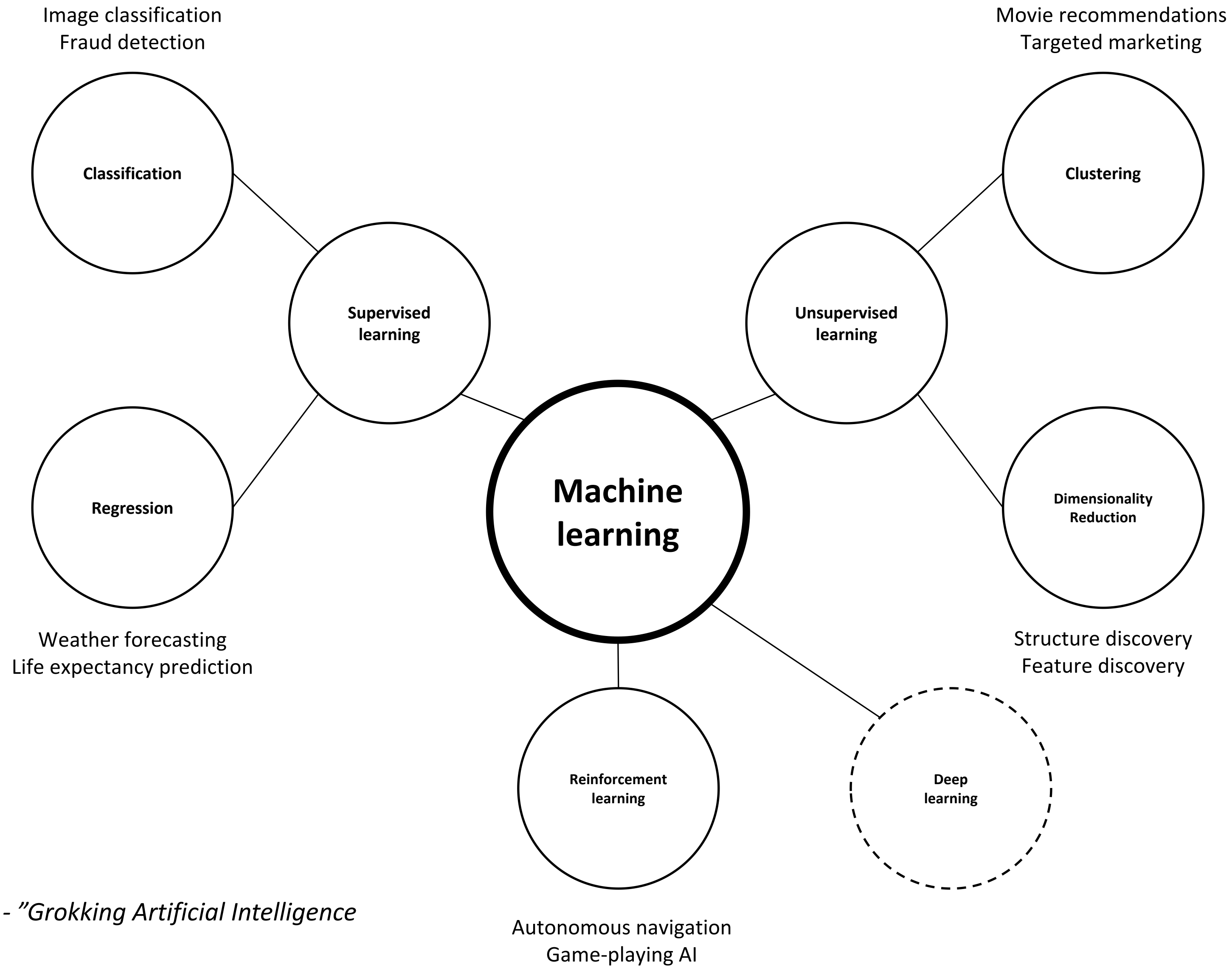
Traditional programming:

Input data + Program (Rules) = Output

Machine learning:

Input data + Output = Program (Rules)

Machine Learning Categories



Based on:
Rishal Hurbans - "Grokking Artificial Intelligence Algorithms"

Main Machine Learning Categories

Supervised learning	Unsupervised learning	Reinforcement learning
Supervised learning is one of the most common techniques in machine learning. It is based on **known relationship(s) and patterns within data** (for example: relationship between inputs and outputs). Frequently used types: **regression**, and **classification**.	**Unsupervised learning** involves finding underlying patterns within data. Typically used in **clustering** data points (similar customers, etc.)	Reinforcement learning is inspired by behavioral psychology. It is **based on a rewarding / punishing an algorithm**. Rewards and punishments are based on algorithm's action within its environment.

What Kind of Questions ML Answers?

Question	ML Category	Example
Is this A or B?	Classification	Will this car fail in the next two months? Yes or no?
Is this weird?	Anomaly detection	Is this credit card charge normal?
How much / many?	Regression	What will the temperature be tomorrow?
How is this organized?	Clustering	Which car models have the most brake problems?
What should I do next?	Reinforcement learning	Adjust room humidity or leave as is?

Supervised Learning: Regression

What is Regression?

Definition:

A technique for **estimating the relationship between a dependent variable** ("outcome") **and one or more independent variables** ("predictors" or "**features**"). The most common form is **linear regression**, in which one **finds the line** (or a more complex linear combination) **that most closely fits the data** (for example using the least-squares method).

Source: https://en.wikipedia.org/wiki/Regression_analysis

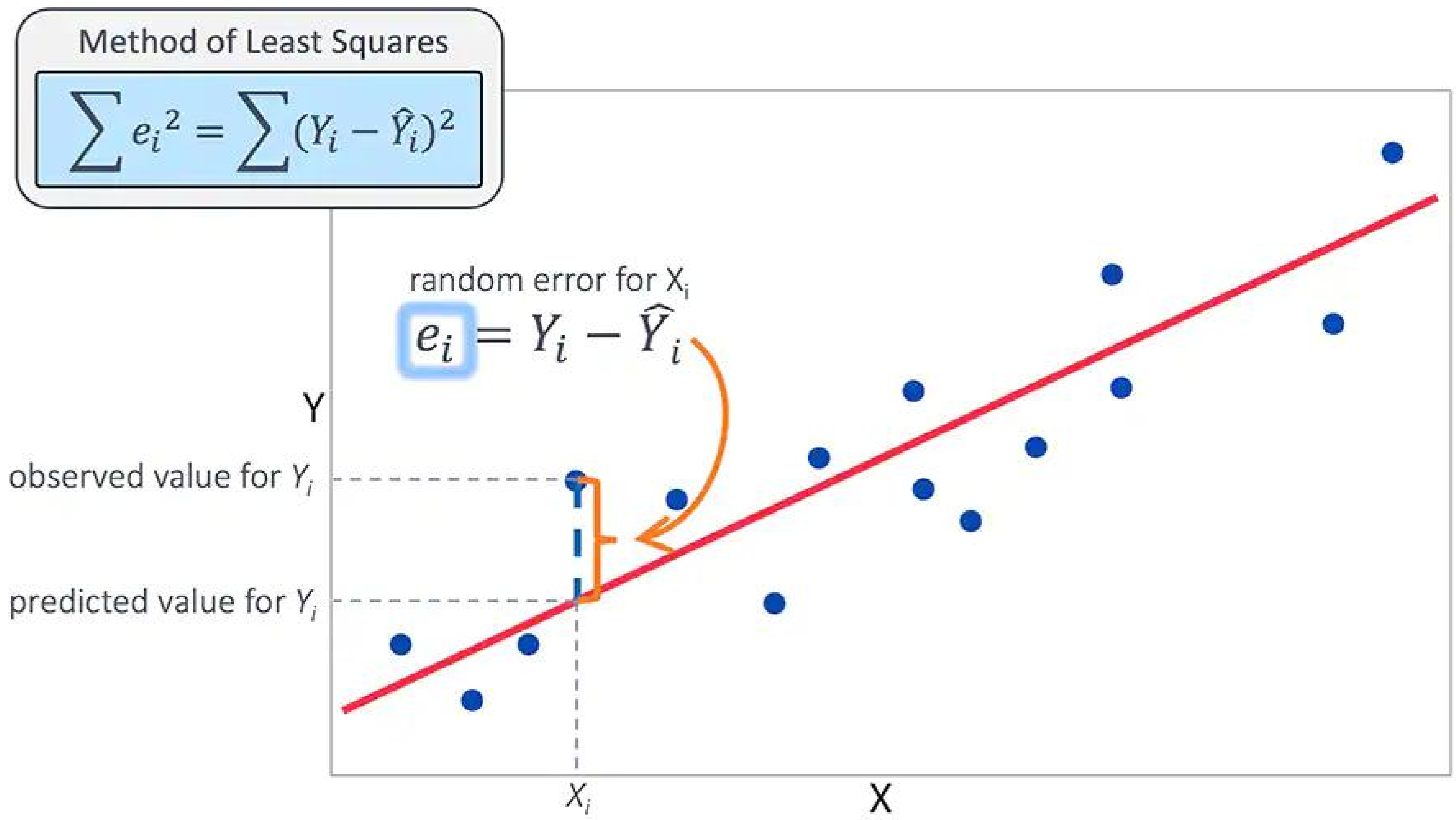
Wait! Why is it called regression?

Origins of ‘Regression’ Term

Source: https://en.wikipedia.org/wiki/Francis_Galton

Sir Francis Galton, an English polymath studied, among other things, heredity in humans. In one experiment he compared children height to their parent heights. He observed that children heights **regressed** towards the average height of an adult.

Linear Regression Using Least-Squares

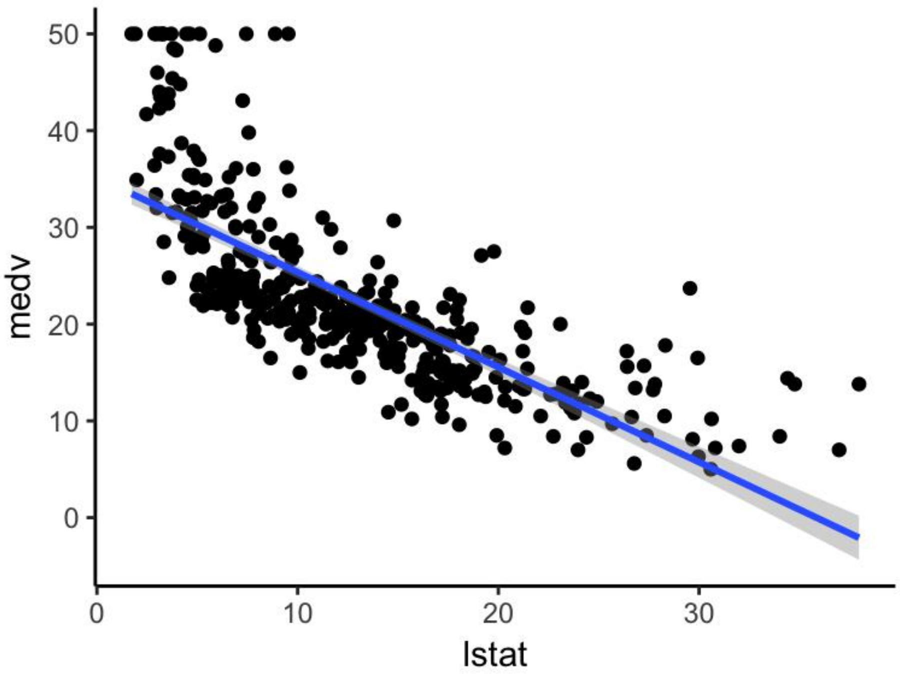


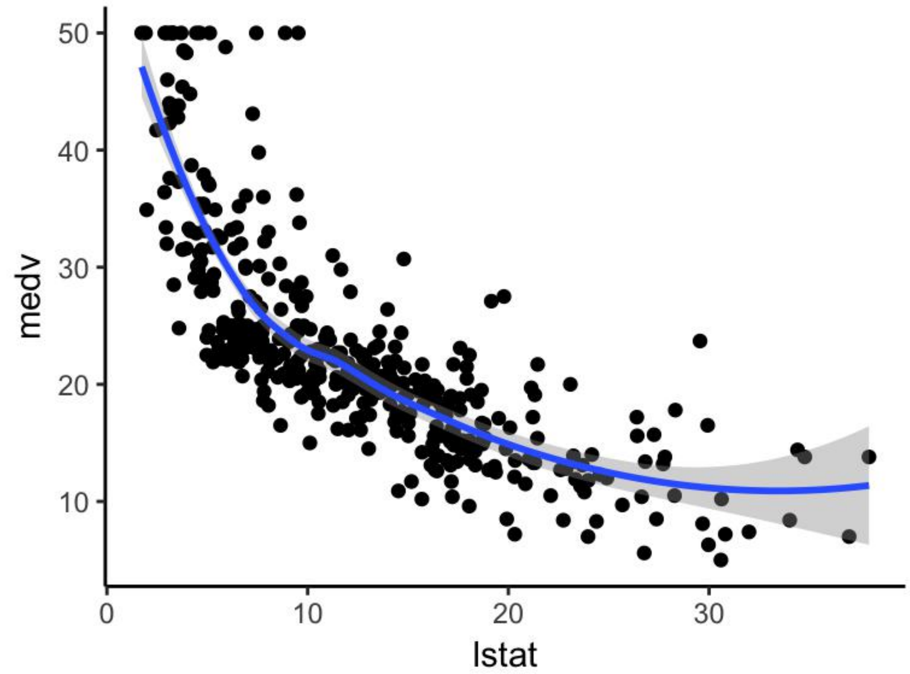
The goal is to find the line y = ax + b that minimizes the amount of error.

Source: https://www.jmp.com/en_us/statistics-knowledge-portal/what-is-multiple-regression/fitting-multiple-regression-model.html

Linear and Nonlinear Regression

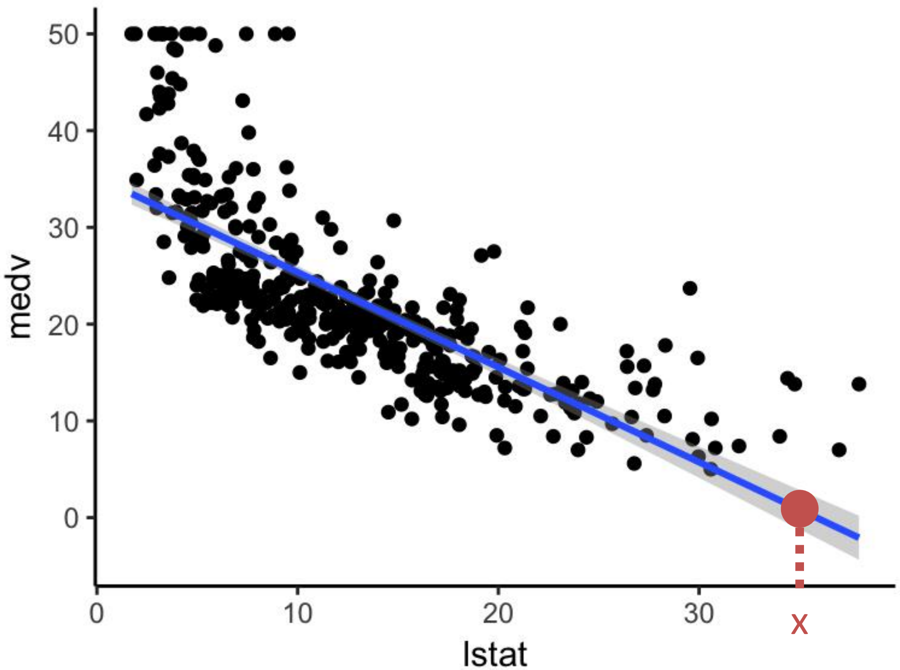
Frequently a polynomial curve will be a better fit to input data than a straight line. A nonlinear regression can be applied as well.

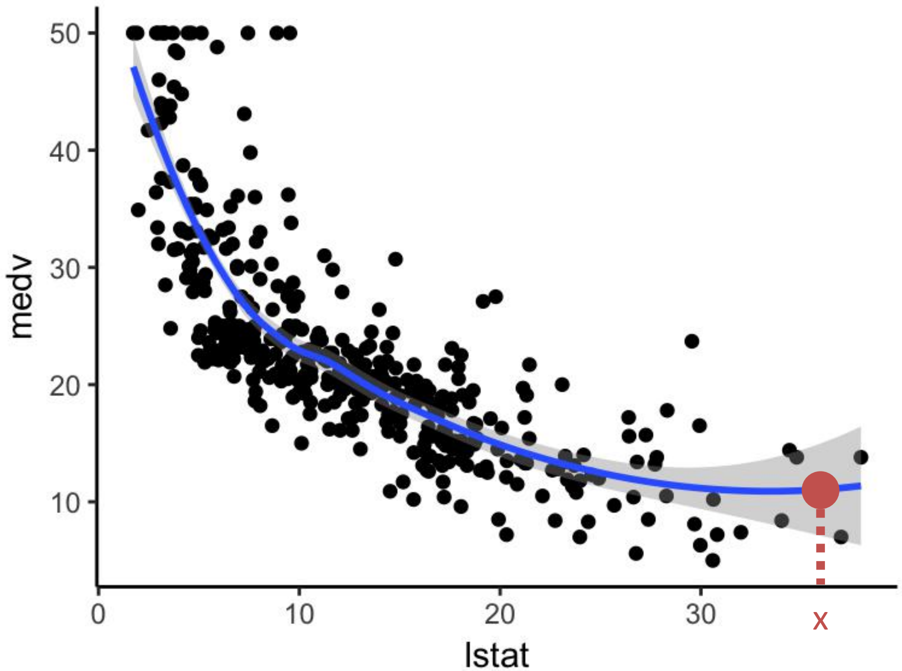




Source: http://www.sthda.com/english/articles/40-regression-analysis/162-nonlinear-regression-essentials-in-r-polynomial-and-spline-regression-models/

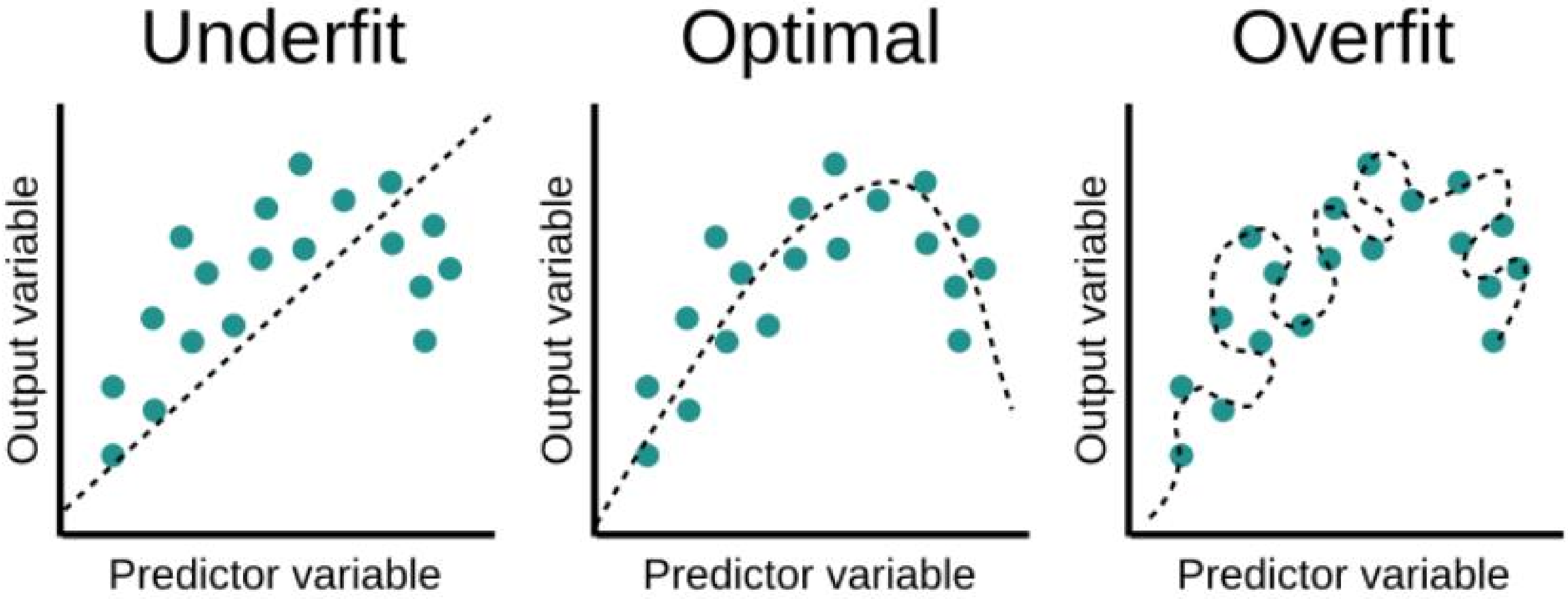
Regression: Prediction / Forecasting





Source: http://www.sthda.com/english/articles/40-regression-analysis/162-nonlinear-regression-essentials-in-r-polynomial-and-spline-regression-models/

Regression: Underfitting / Overfitting



Source: https://livebook.manning.com/book/machine-learning-for-mortals-mere-and-otherwise/chapter-9/v-4/29

Supervised Learning: Classification

What is Classification?

Definition:

Classification is a process of **categorizing data into distinct classes**. In practice it means **developing a model that maps input data to a discrete set of labels / targets**. Classification can be:

- **binary** - there is only two classes: yes / no, true / false, spam / not spam
- **multi-class** - there are multiple classes available, only one is assigned
- **multi-label** - multiple classes an be assigned

Exercise 1: Regression or Classification?

Ex. 1: Regression or Classification?

1. Based on data about rats: we have a life-expectancy variable and obesity variable. We are trying to find a correlation between the two.

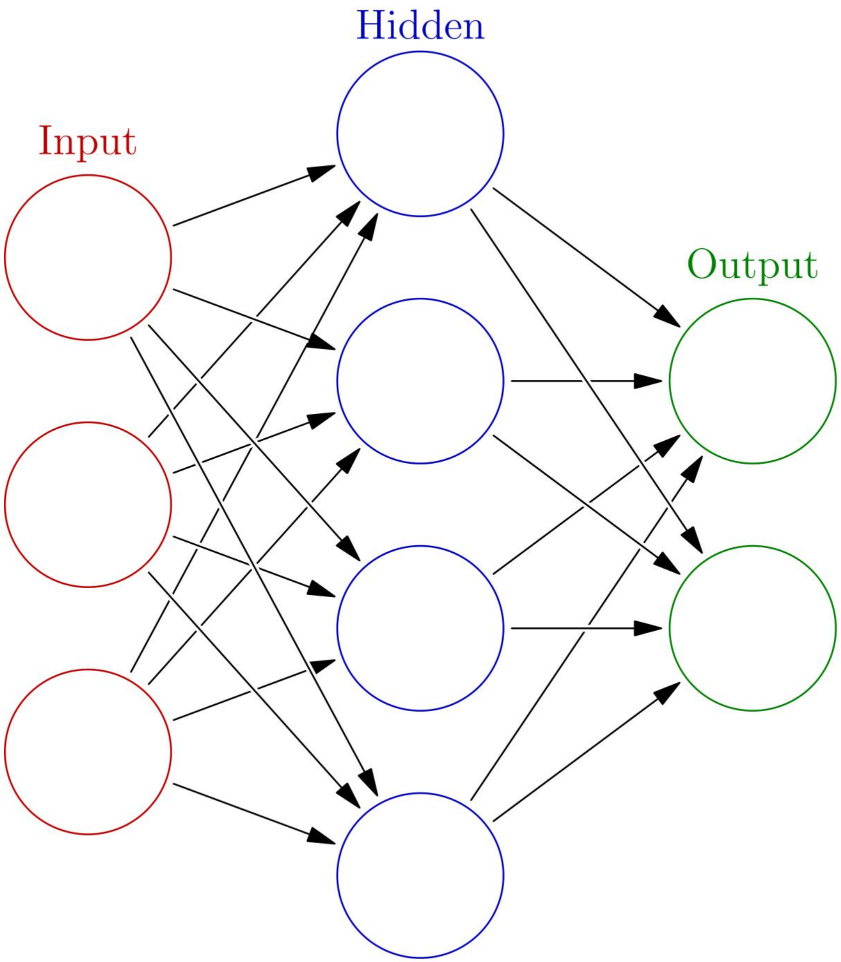
2. Based on data about animals: we have weight for each animal and information about whether it has wings or not. We are trying to determine which animals are birds?

3. Based on data about computing devices: we have the screen size, weight, and operating system of several devices. We want to determine which devices are tablets, laptops or phones.

4. Based on data about weather: we have the amount of rainfall and a humidity value. We want to determine humidity in different rainfall seasons.

Classification with Artificial Neural Networks

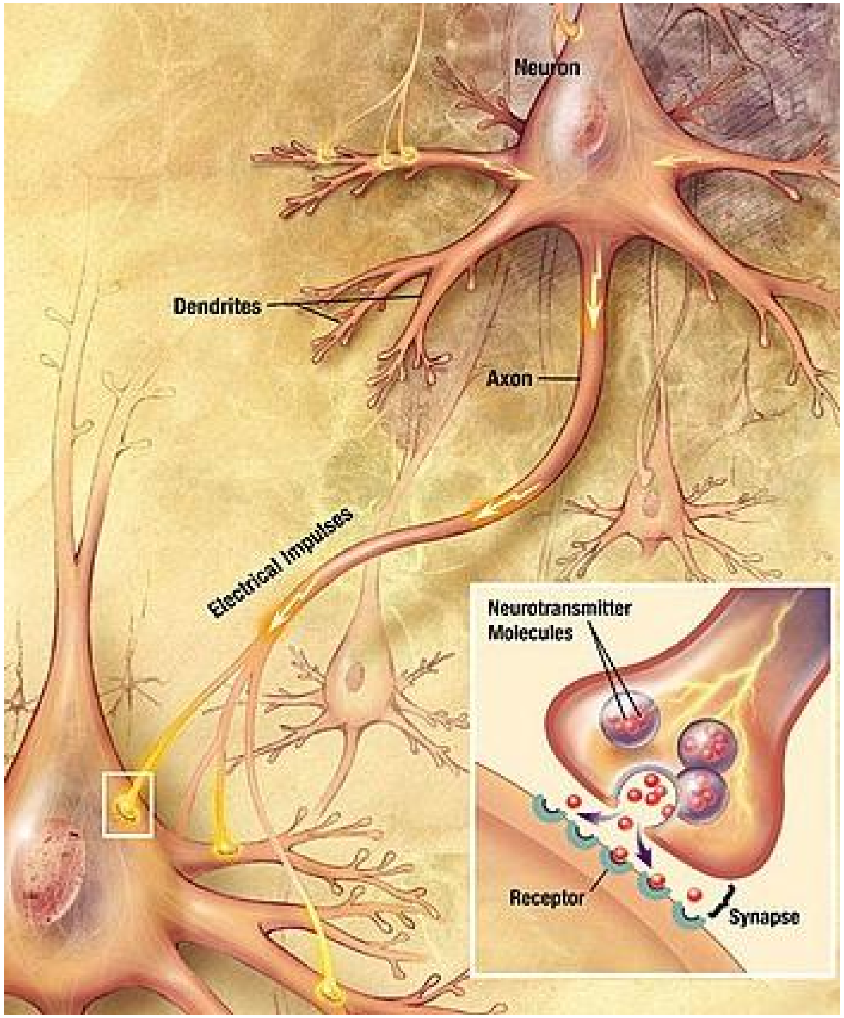
Artificial Neural Network (1943)



Source: https://en.wikipedia.org/wiki/Artificial_neural_network

First computational models of an **Artificial Neural Network** (loosely inspired by biological neural networks) were proposed by Warren McCulloch and Walter Pitts in 1943. Their ideas are a **key component of modern day machine and deep learning**.

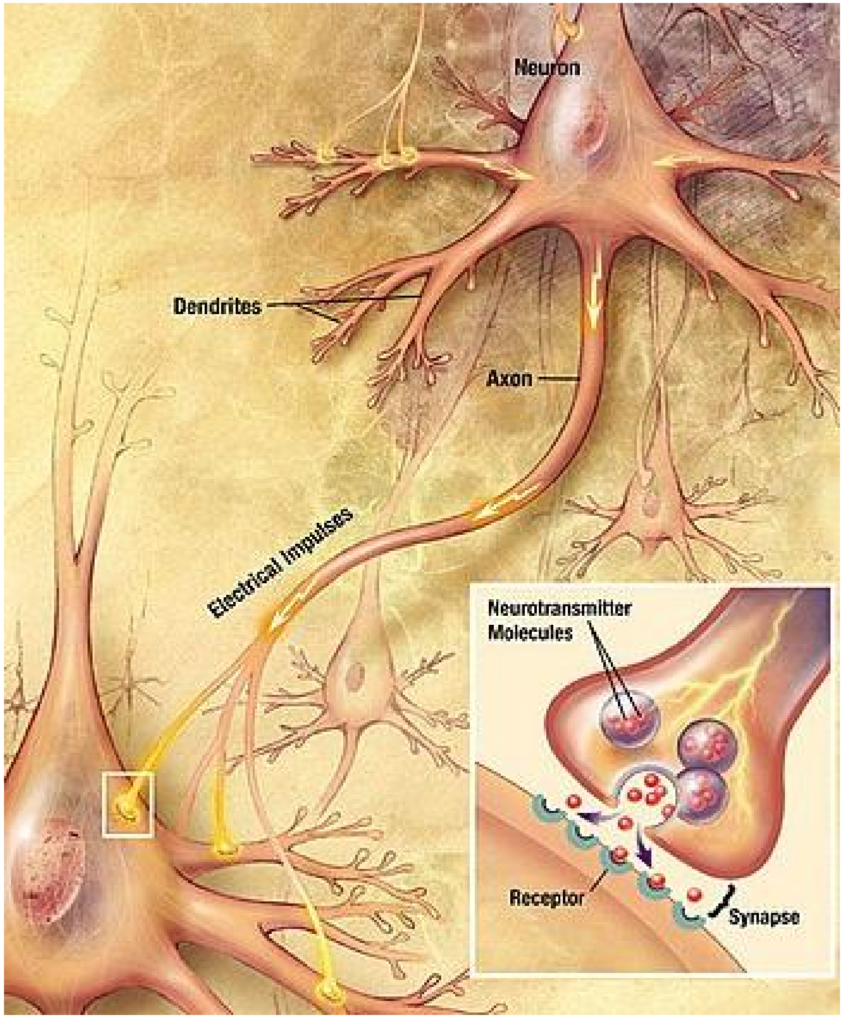
A Biological Neuron



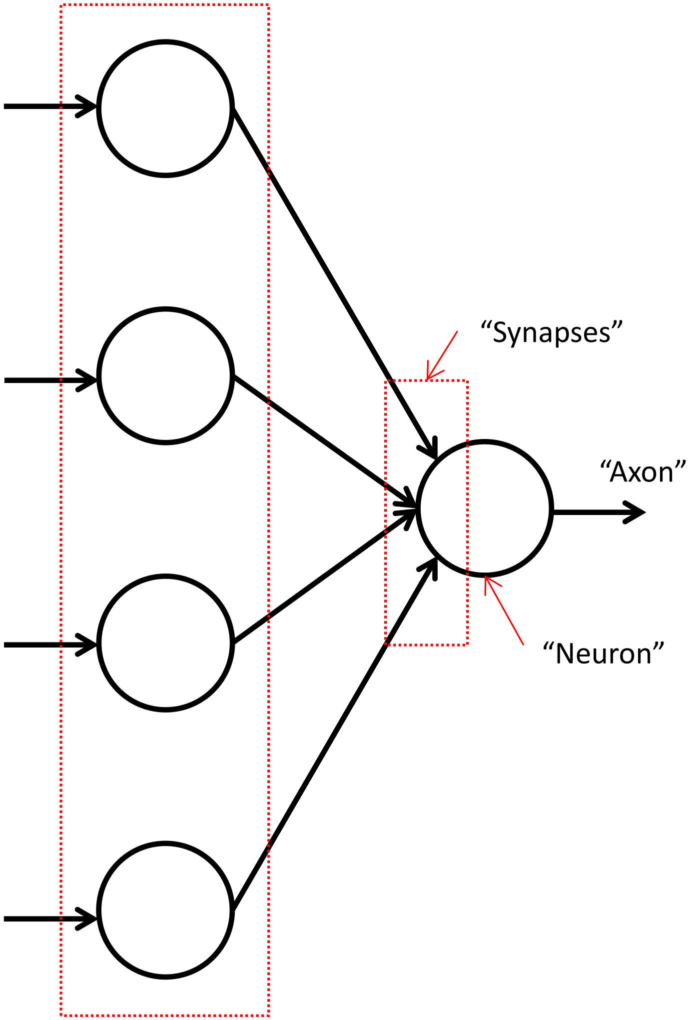
Source: https://en.wikipedia.org/wiki/Neuron

A **neuron** or nerve cell is an electrically excitable cell that **communicates with other cells via specialized connections called synapses**. Most **neurons receive signals via the dendrites and soma** and **send out signals down the axon**. At the majority of synapses, signals cross from the axon of one neuron to a dendrite of another.

Biological vs. Artificial Neuron



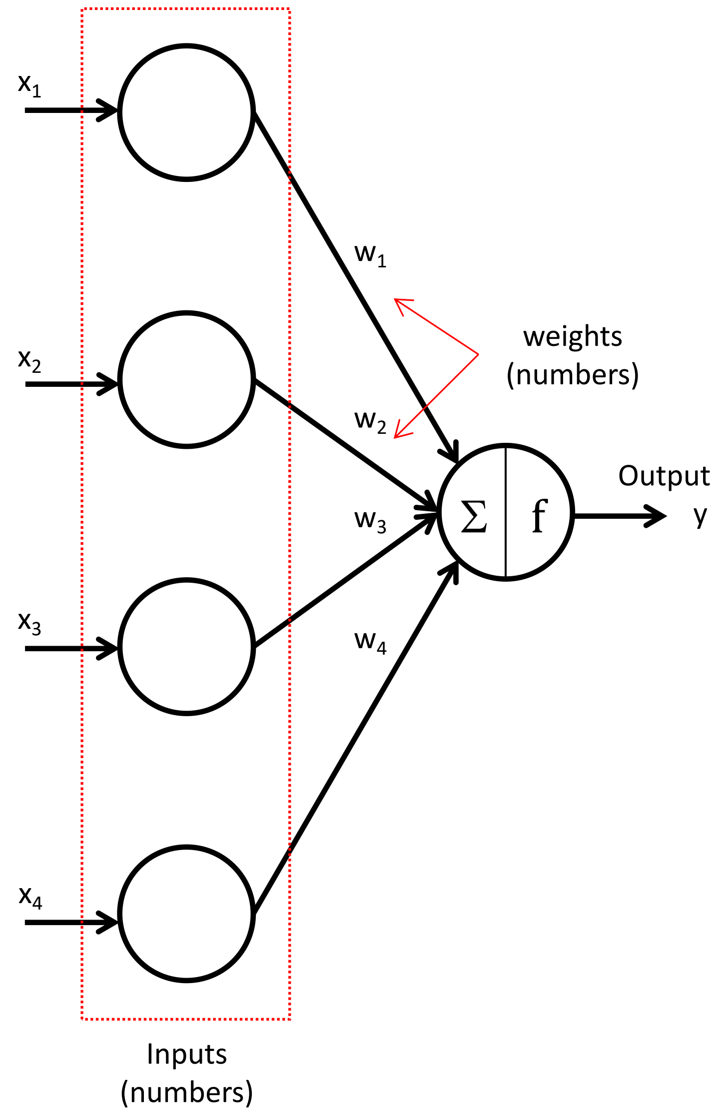
Source: https://en.wikipedia.org/wiki/Neuron



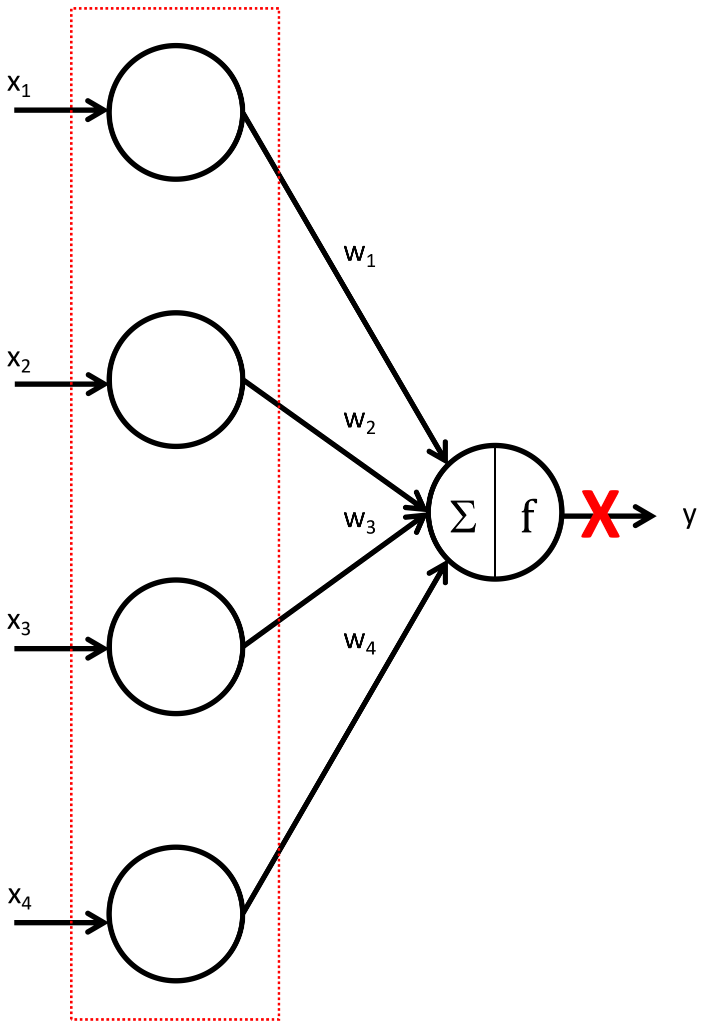
Artificial Neuron (Perceptron)

A (**single-layer**) **perceptron** is a model of a biological neuron. It is made of the following components:

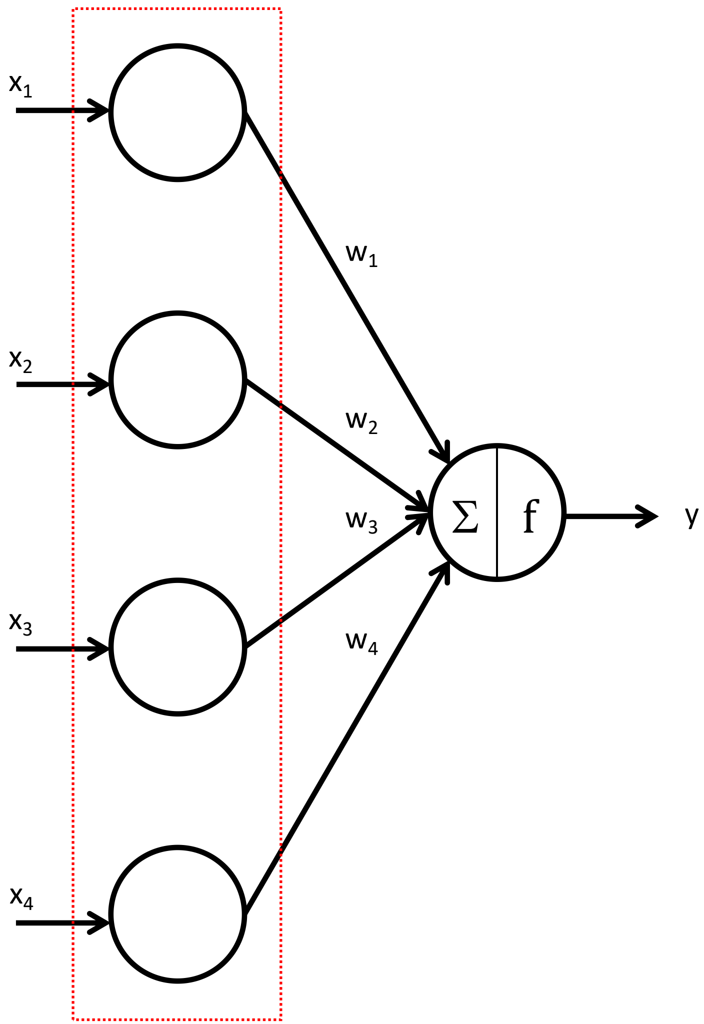
- inputs x_i - numerical values representing information
- weights w_i - numerical values representing how "important" corresponding input is
- weighted sum: $\Sigma\, w_i * x_i$
- activation function f that decides if the neuron "fires"



Artificial Neuron (Perceptron)

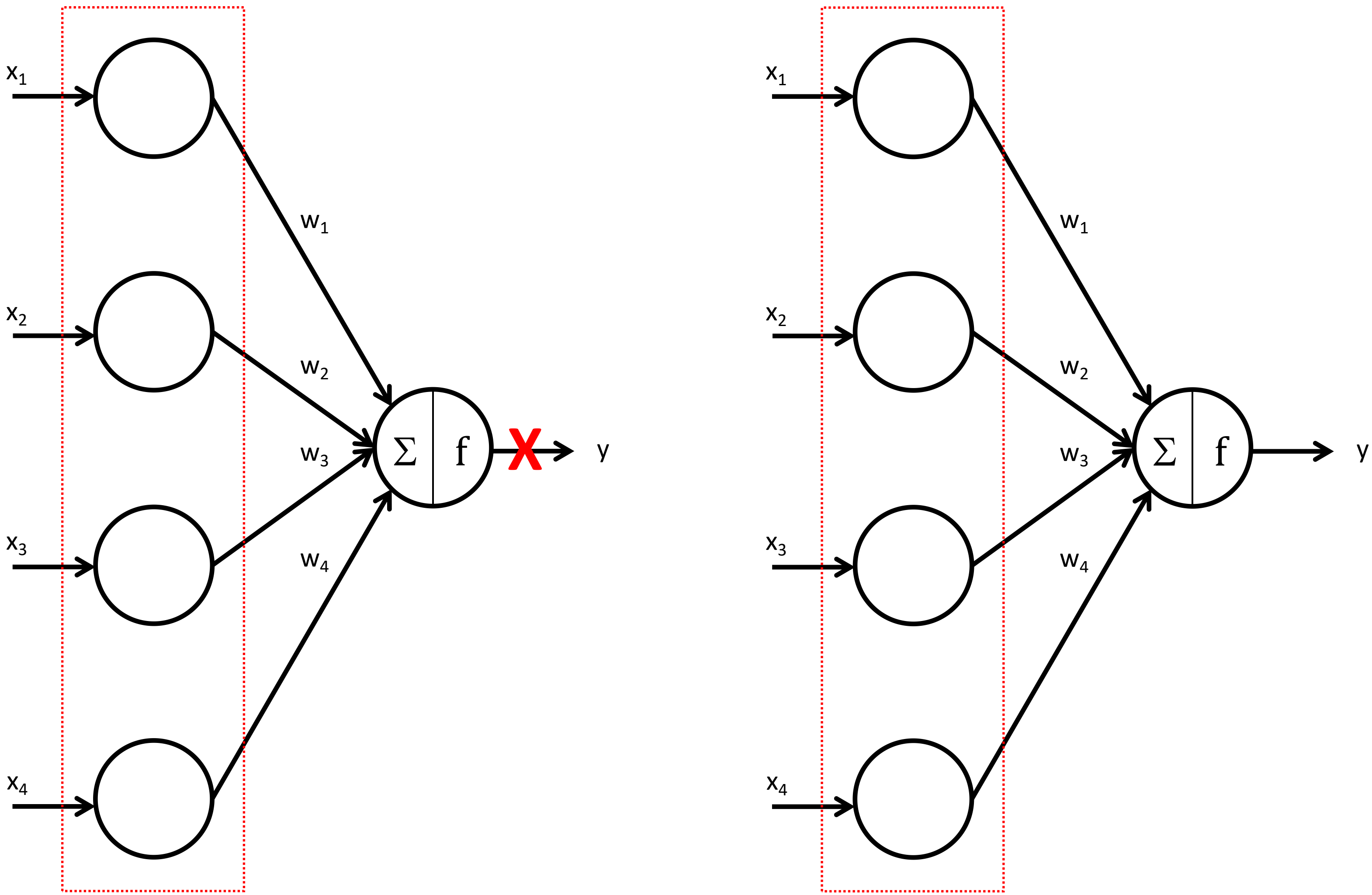


$\Sigma\, w_i * x_i < 0 \rightarrow f = 0$ → **DON'T "fire"**



$\Sigma\, w_i * x_i \geq 0 \rightarrow f = 1$ → **"fire"**

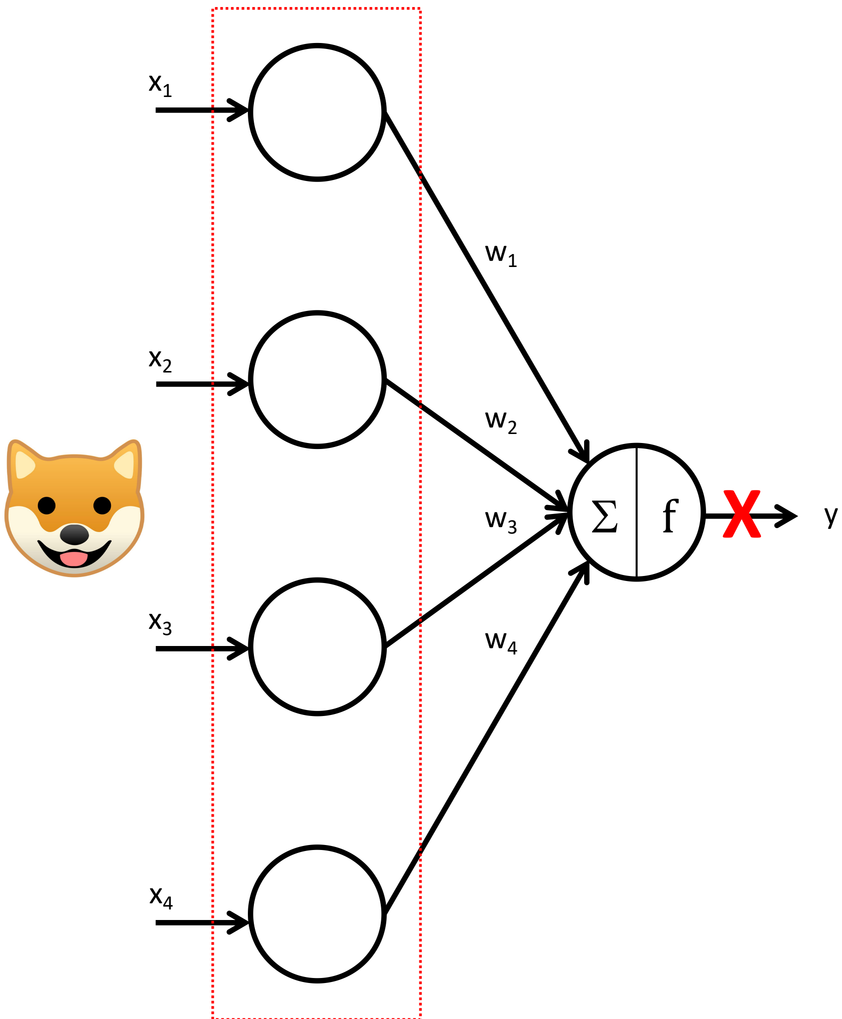
Single-layer Perceptron as a Classifier



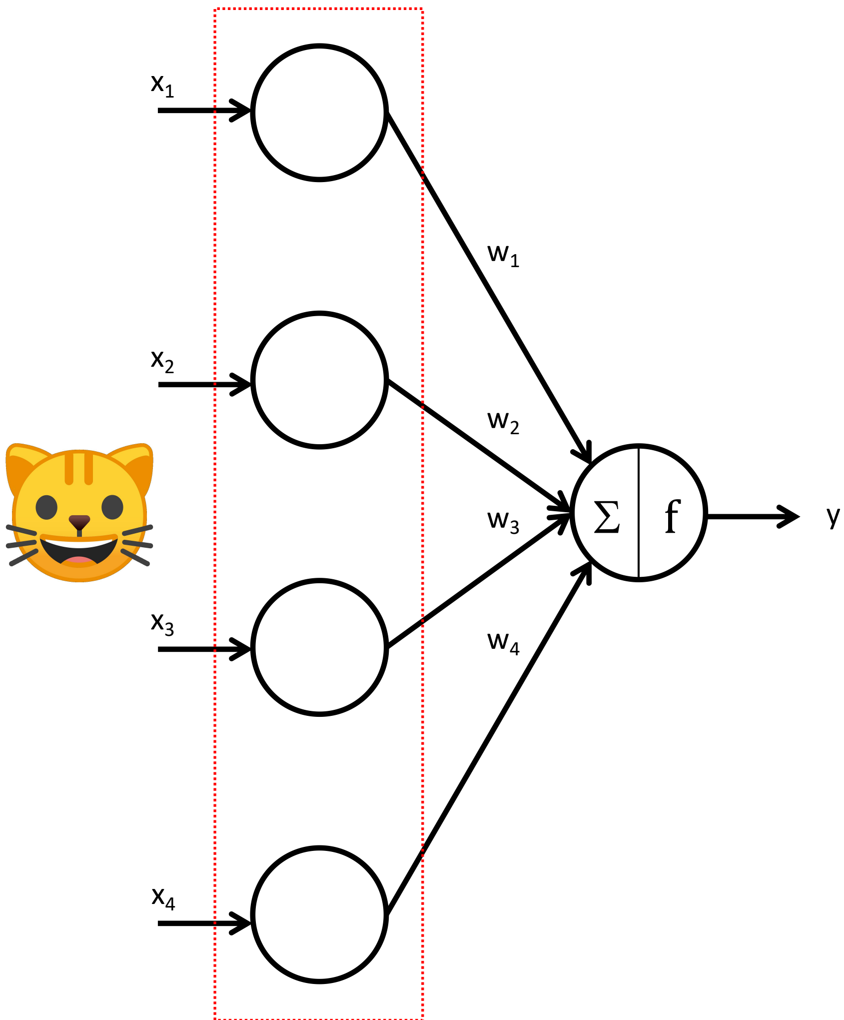
$\Sigma\, w_i * x_i < 0 \rightarrow f = 0$ → **NO**

$\Sigma\, w_i * x_i \geq 0 \rightarrow f = 1$ → **YES**

Single-layer Perceptron as a Classifier



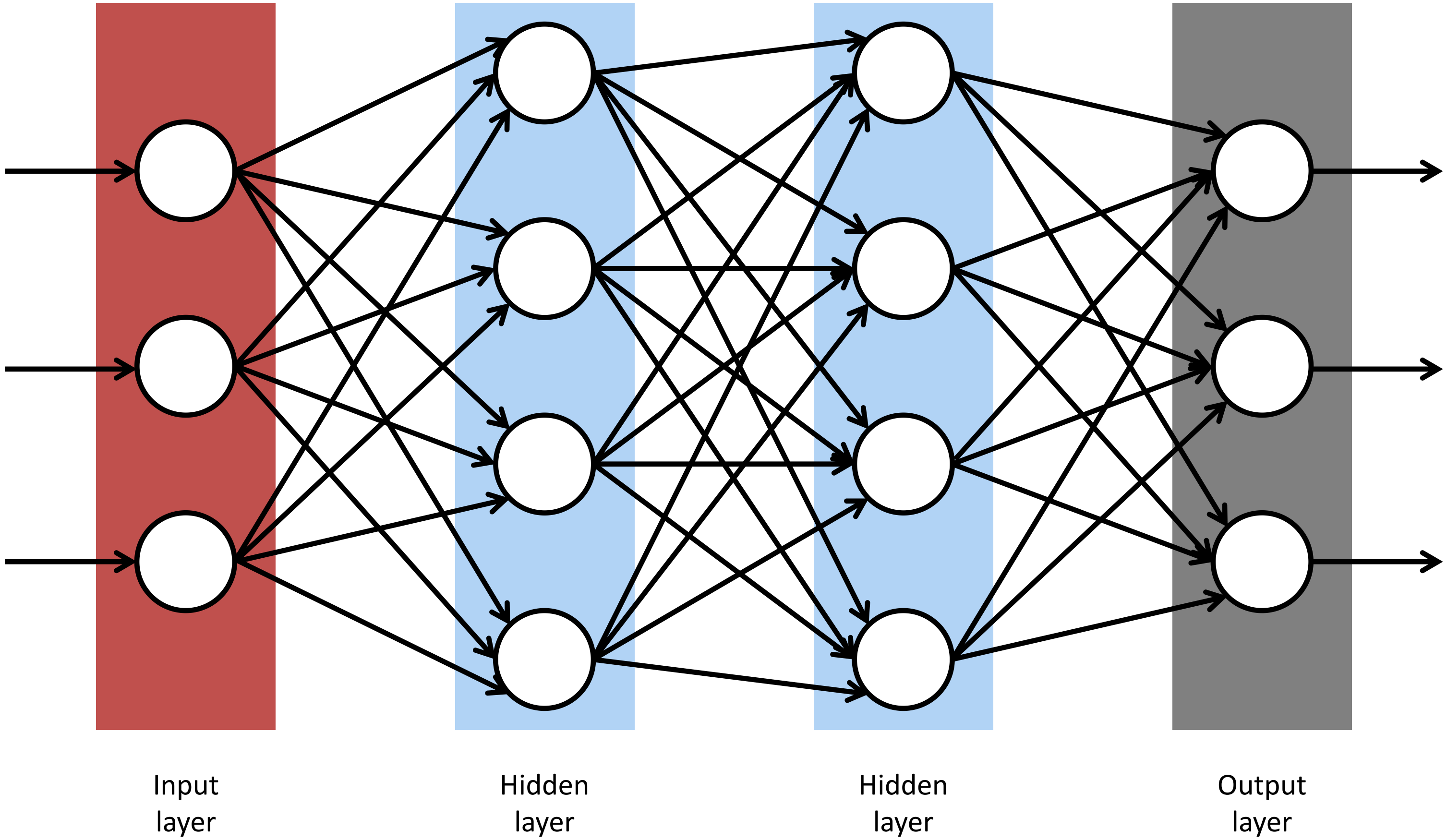
$\Sigma\, w_i * x_i < 0 \rightarrow f = 0$ → **NOT CAT**



$\Sigma\, w_i * x_i \geq 0 \rightarrow f = 1$ → **CAT**

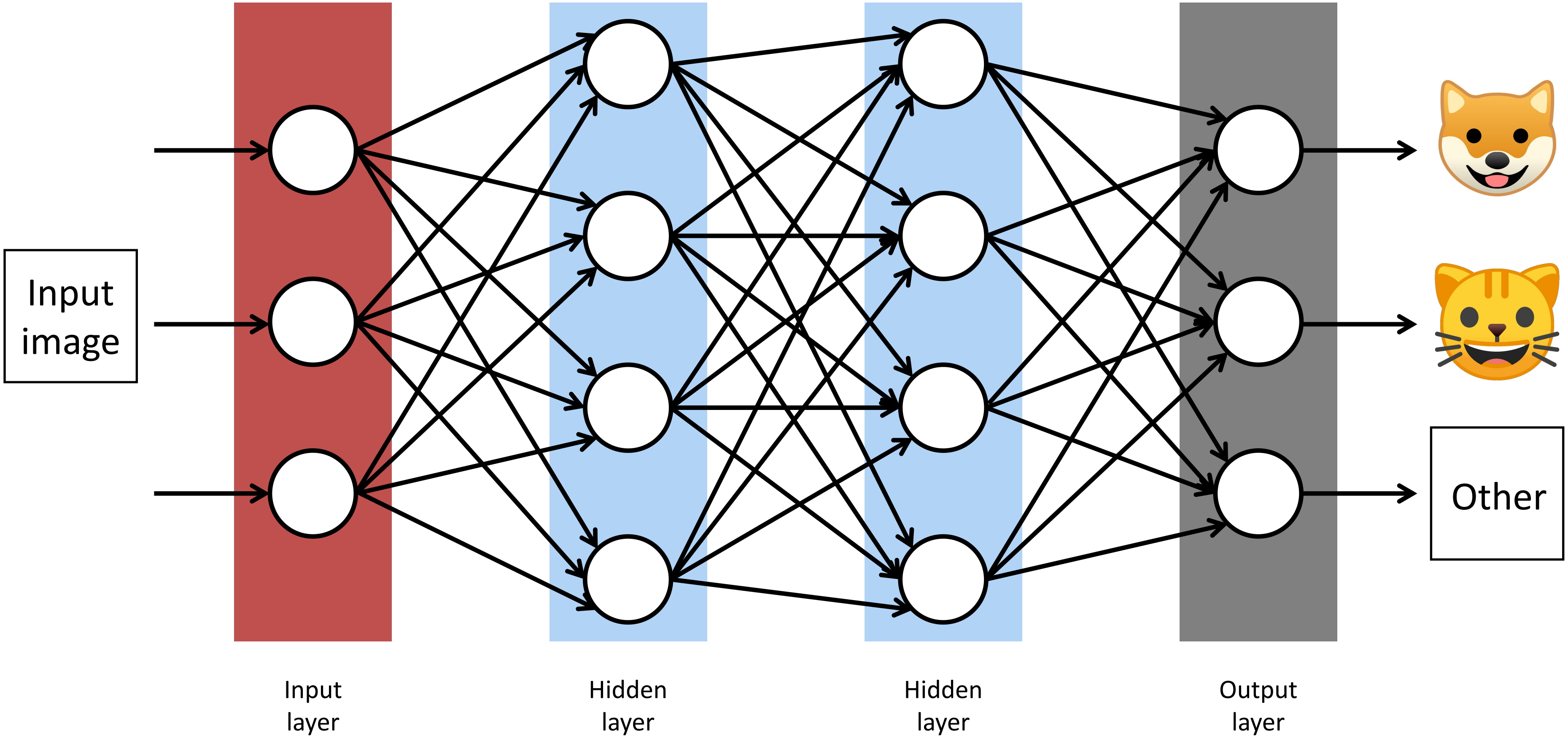
Artificial Neural Network (ANN)

An artificial neural network is made of **multiple artificial neuron layers**.

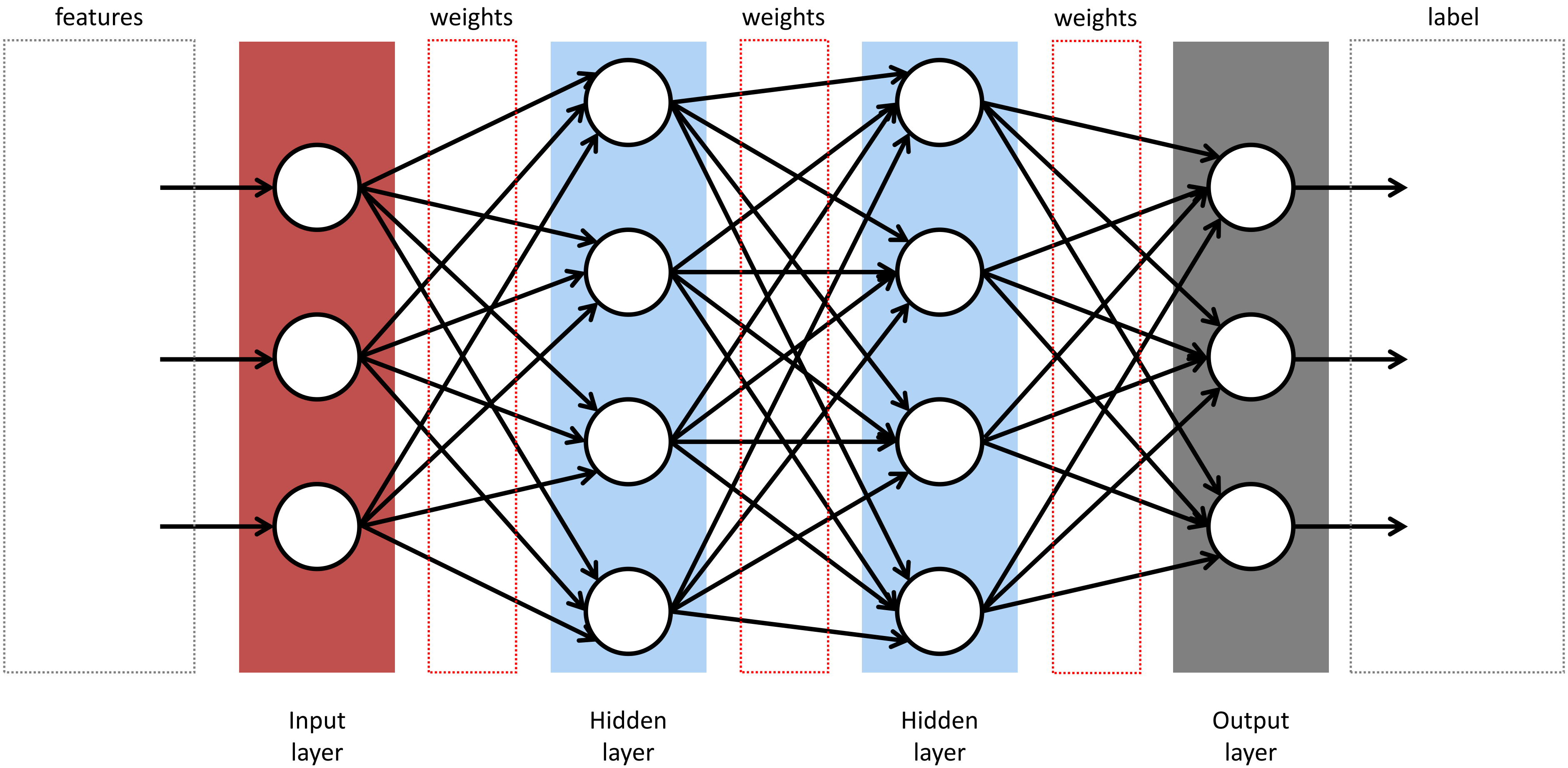


ANN as an Image Classifier

An artificial neural network can be used as a **classifier** as well.

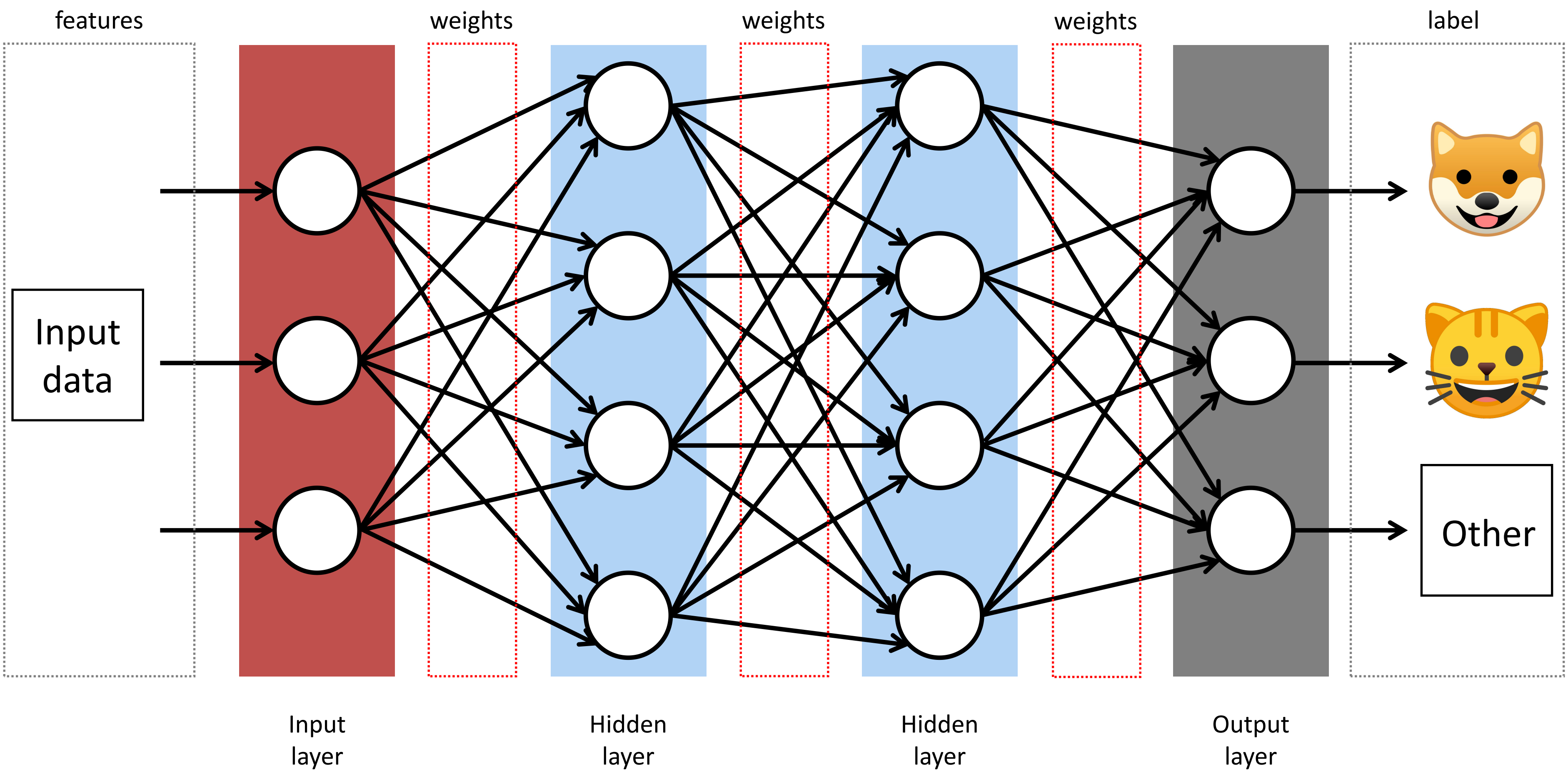


ANN as a Classifier



ANN: Supervised Learning

In order to work properly a classifier **needs to be trained** first with **labeled data**.



Training will **adjust all the weights** within this artificial neural network.

Training Data: Features + Labels

Typically input data will be represented by a **limited set of features**.

Features:
Wheels: 4
Weight: 8 tons
Passengers: 1

Label:
Truck

Features:
Wheels: 6
Weight: 8 tons
Passengers: 1

Label:
Truck

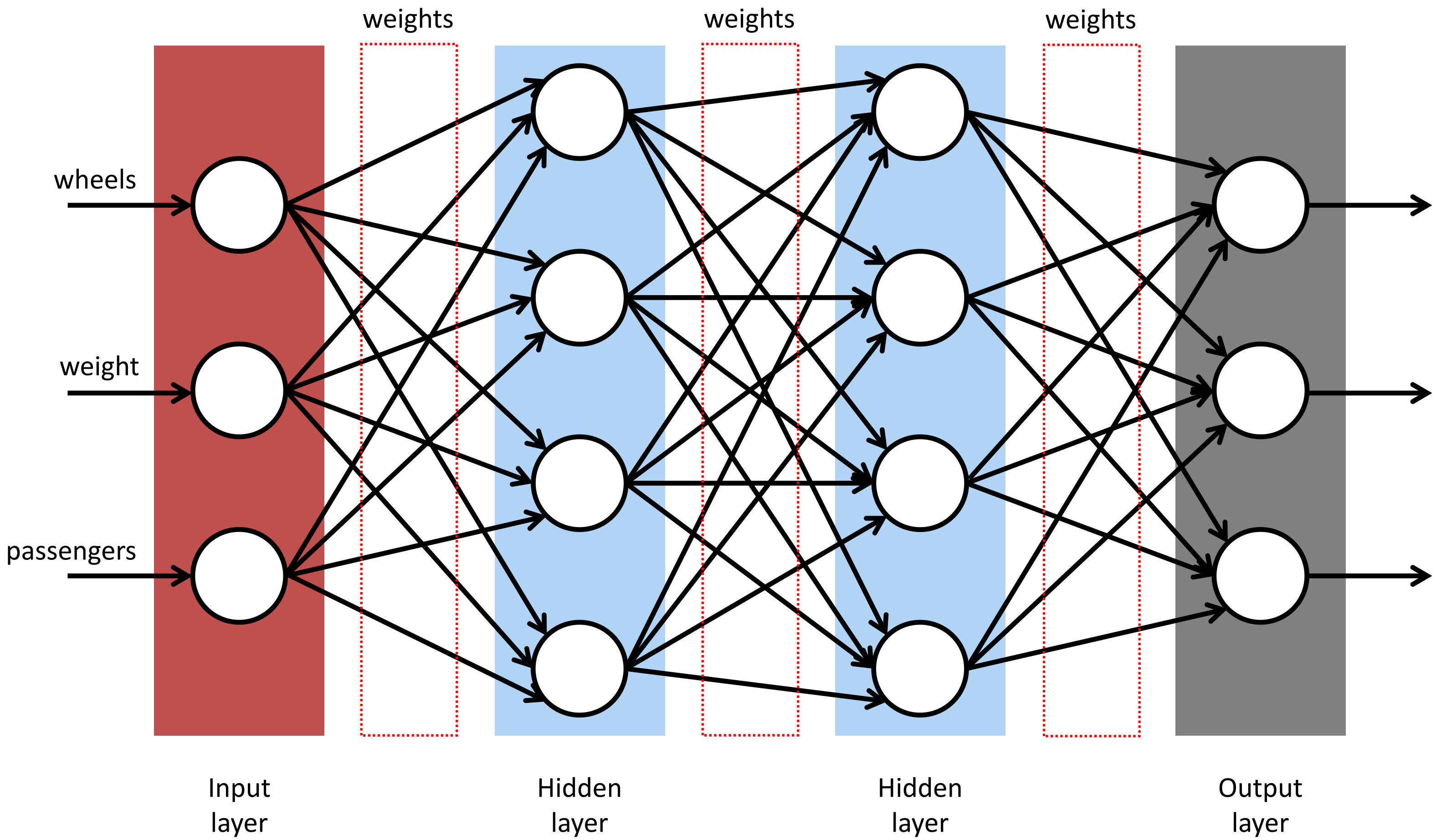
Features:
Wheels: 4
Weight: 1 ton
Passengers: 4

Label:
Car

Features:
Wheels: 4
Weight: 2 tons
Passengers: 4

Label:
Car

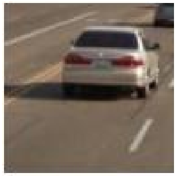
ANN: Supervised Learning



Training Data: Images + Labels

A classifier **needs to be "shown" thousands of labeled examples to learn**.

Label:
BUS



Label:
CAR

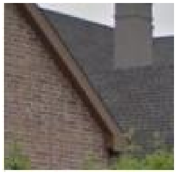
Label:
BRIDGE

Label:
PALM

Label:
TRAFFIC LIGHT

Label:
TAXI

Label:
CROSSWALK



Label:
CHIMNEY

Label:
MOTORCYCLE

Label:
STREET SIGN

Label:
HYDRANT

Label:
BICYCLE

Note how some images are "incomplete" and "flawed".

Training Data: Images + Labels

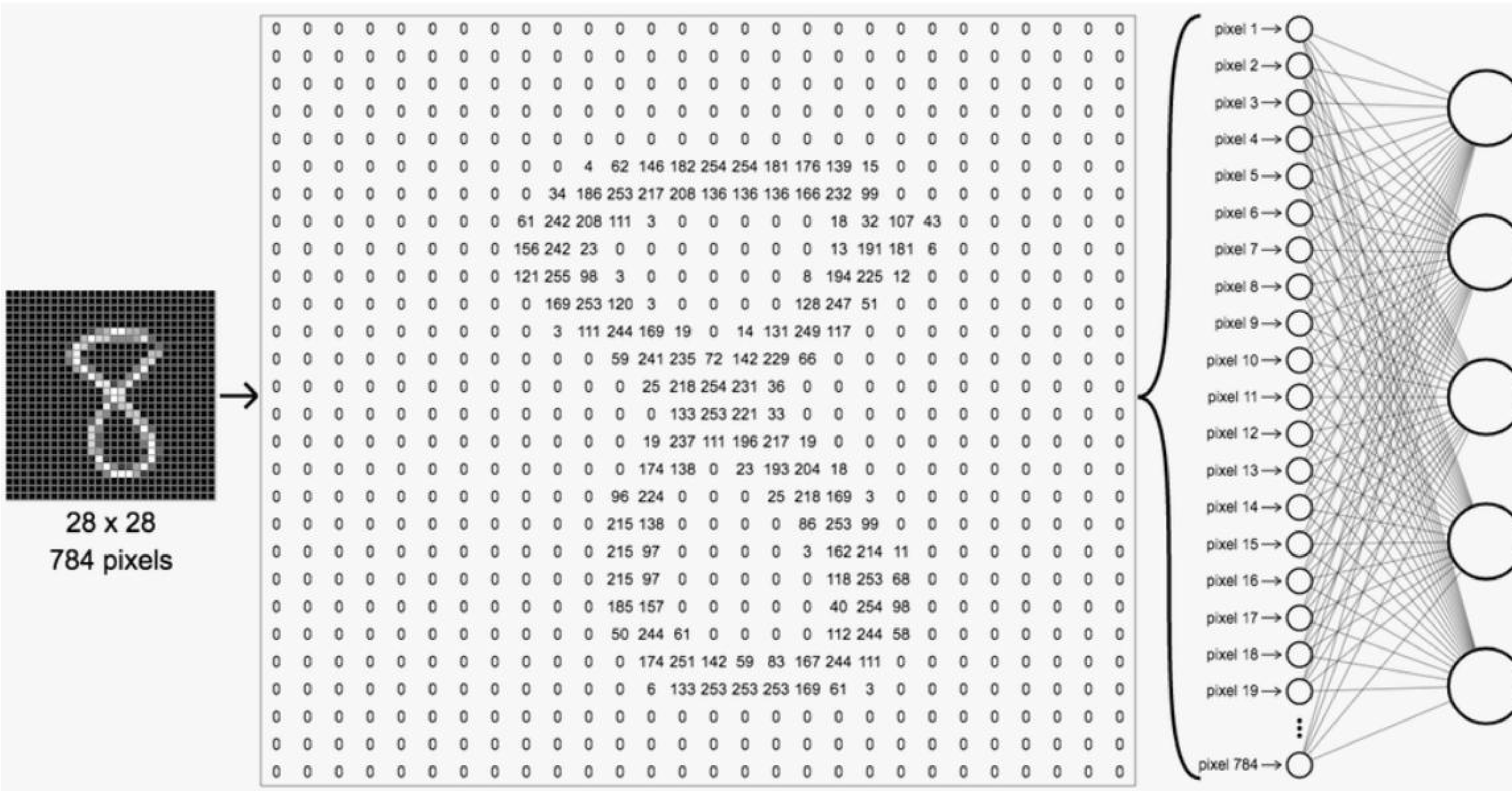
A classifier **needs to be "shown" thousands of labeled examples to learn**.

Source: https://en.wikipedia.org/wiki/MNIST_database

How would you decide which digit it is?

Digit Image as ANN Feature Set

Individual features need to be “extracted” from an image. An image is numbers.



Source: https://nikolanews.com/not-just-introduction-to-convolutional-neural-networks-part-1/

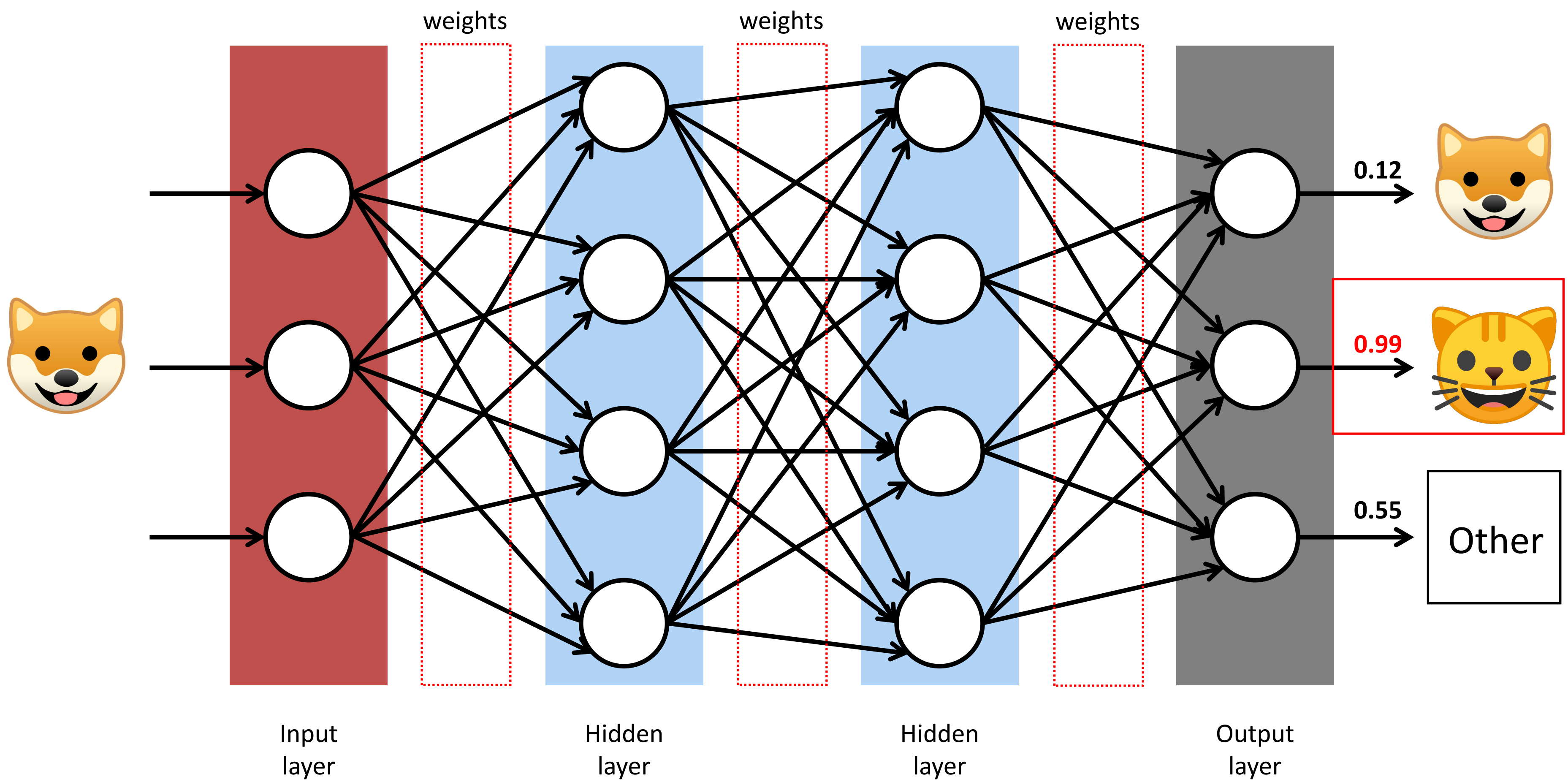
Demo 1: Digit Recognition

http://web-digits-recognizer.herokuapp.com/

https://henryjin.dev/demo/mnist/

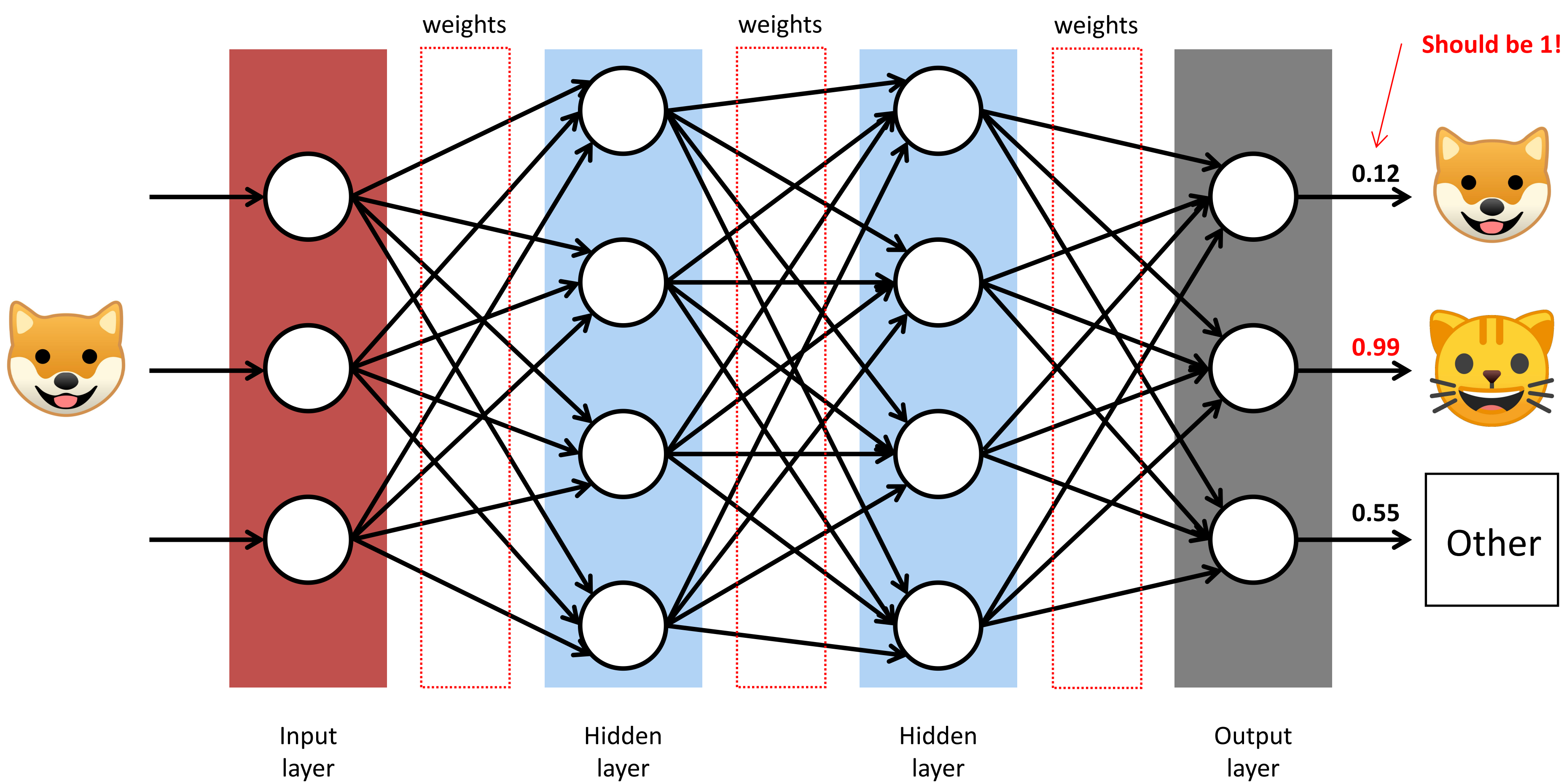
ANN: Supervised Learning

An **untrained classifier** will **NOT label input data correctly**.



ANN: Training

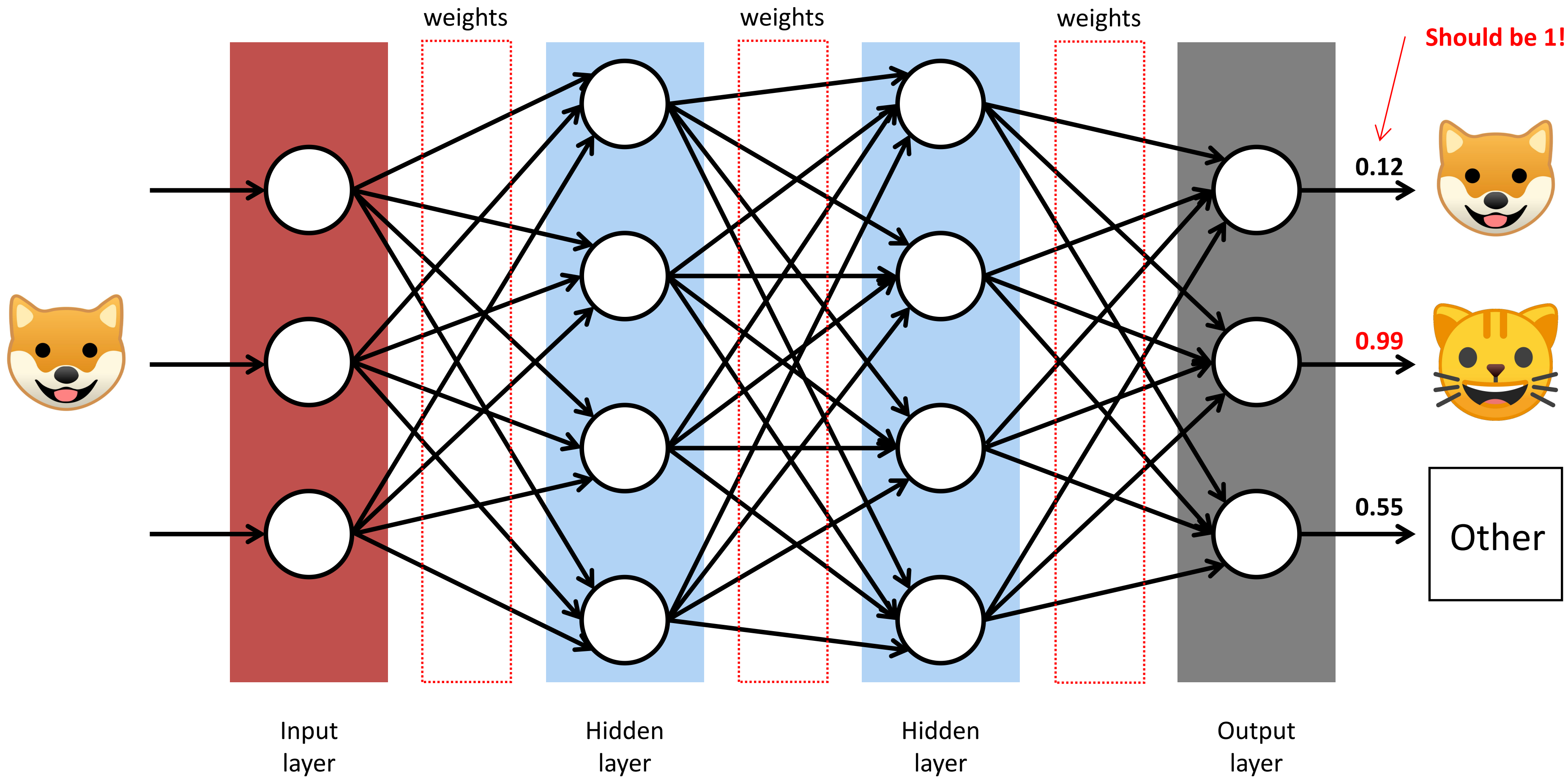
Given: input data and it's corresponding **expected** label: DOG calculate "error".



"Error" = 0.88. Go back and **adjust all the weights** to ensure it is lower next time.

ANN: Training

Show data / label pair: 🐶 / DOG.



Correct all the weights. Repeat many times.

Demo 2: Quick Draw!

https://quickdraw.withgoogle.com/