Machine Learning and Deep Learning

LECTURE 3

Machine Learning and Deep Learning

1. Machine Learning Explained
2. Decision Tree Classifier
3. Deep Learning Ecosystem
4. Resources

Here is a breakdown of AI…

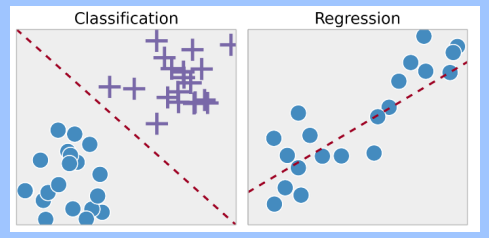
If I say, **Sky**, you would likely say…. **Blue**

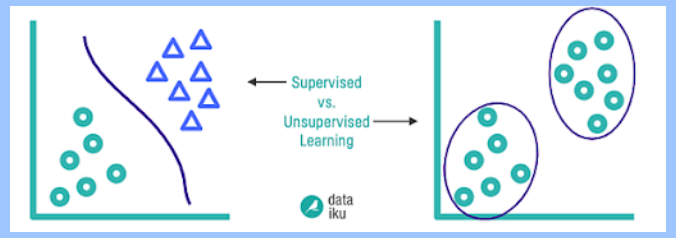
If I say, **grass**, you would likely say…. **Green**

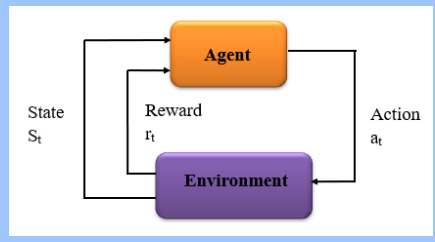
If I say, **AI**, I encourage you to say…. **Machine Learning**

If I say, **machine learning**, I'd like to think… **Prediction**

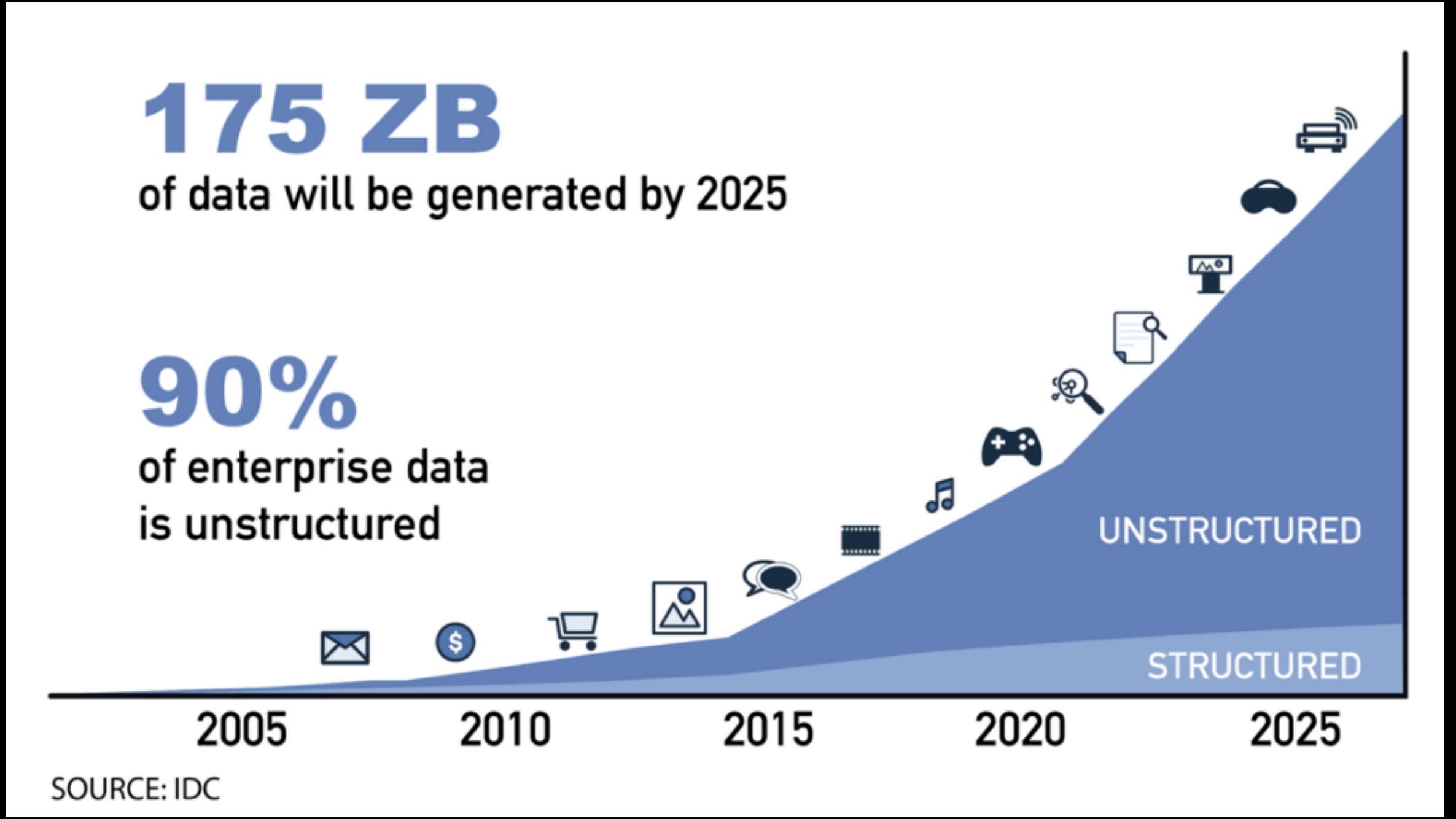
1) Supervised learning **2) Unsupervised learning** **3) Reinforcement learning**







…and what comprises **algorithms'** favorite food???…. **Data**


175 ZB
of data will be generated by 2025
90%
of enterprise data
is unstructured
UNSTRUCTURED
STRUCTURED
2005
2010
2015
2020
2025
SOURCE: IDC

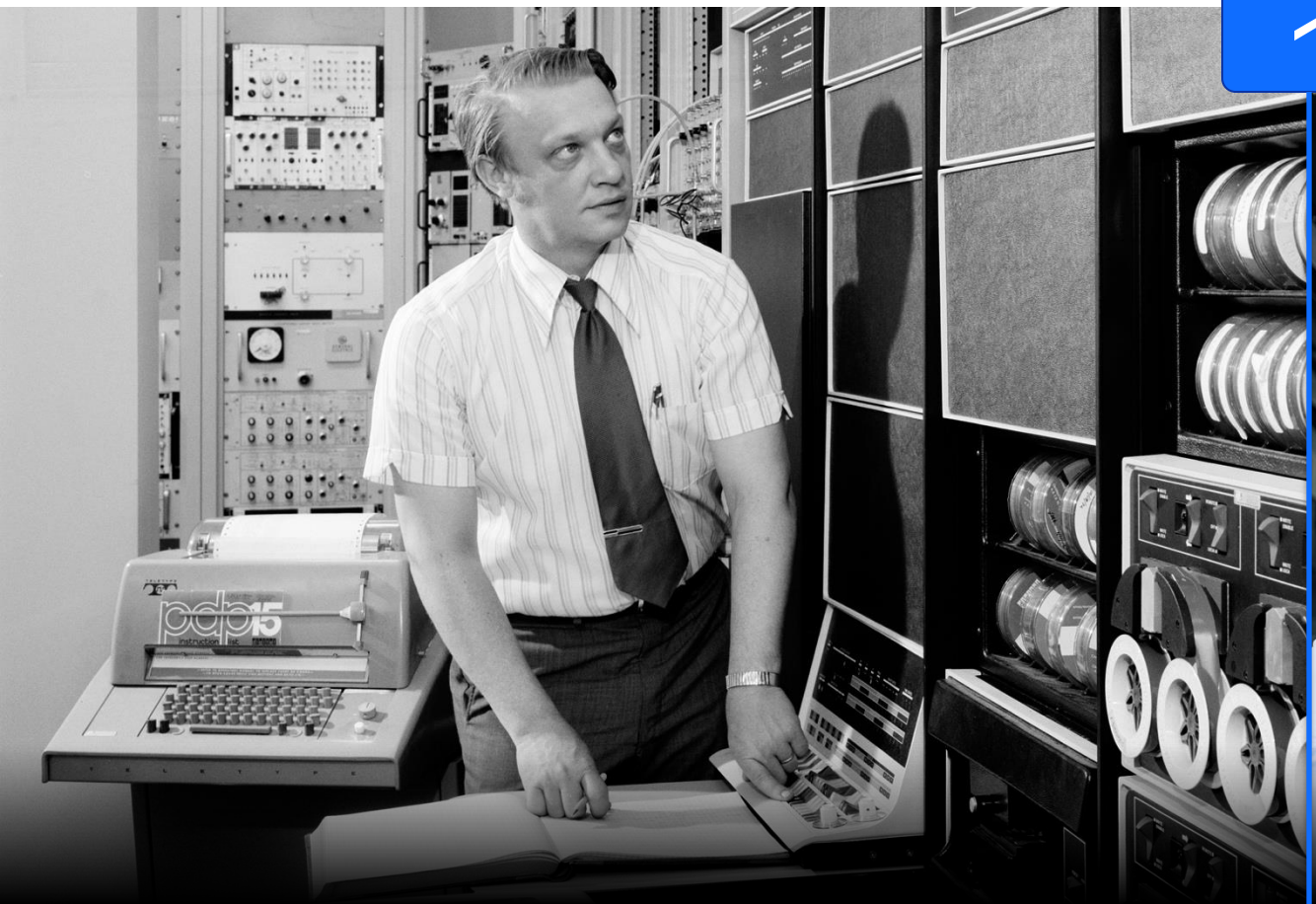
Traditional systems and AI systems

Traditional systems

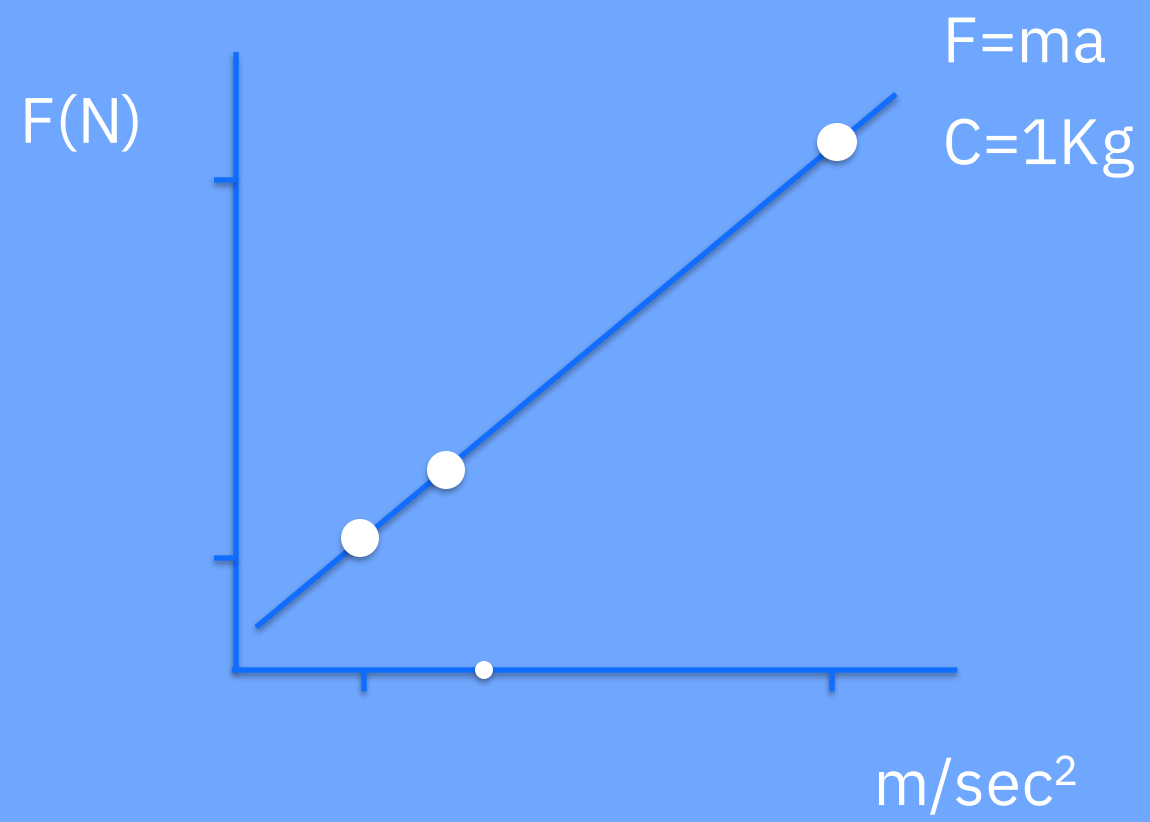
Deterministic
Static

AI systems

Probabilistic
Self-improving

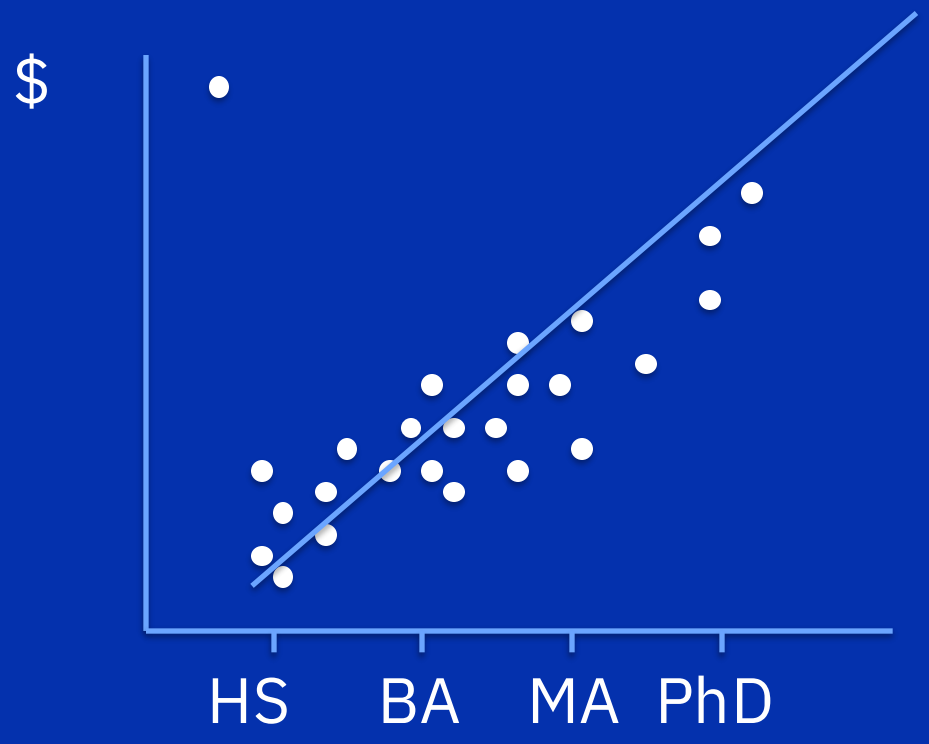


Traditional Deterministic Systems

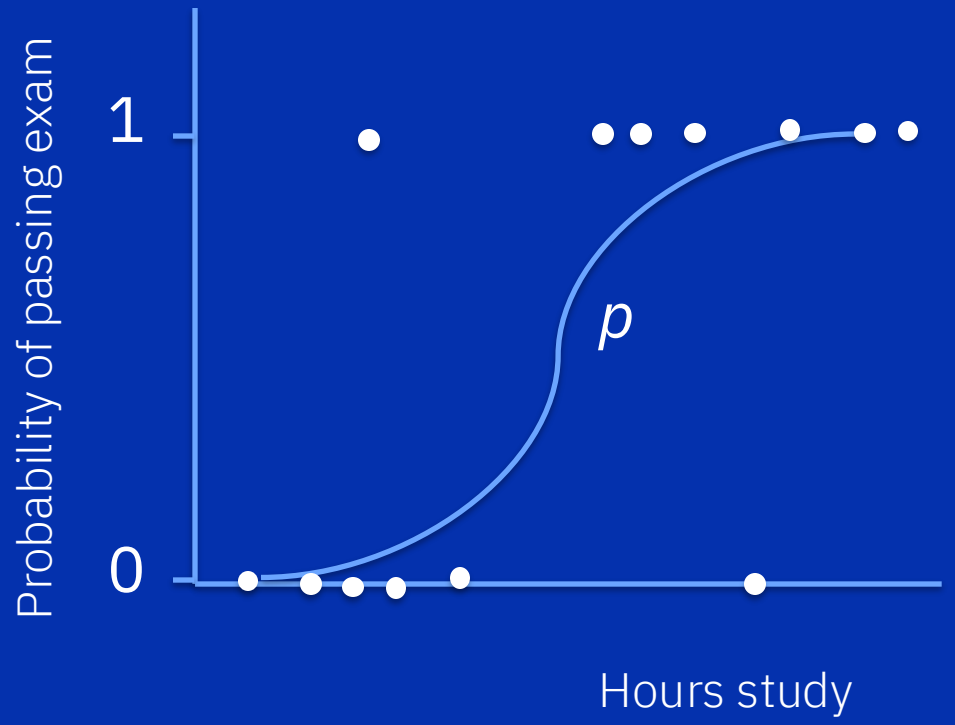


AI Probabilistic Systems

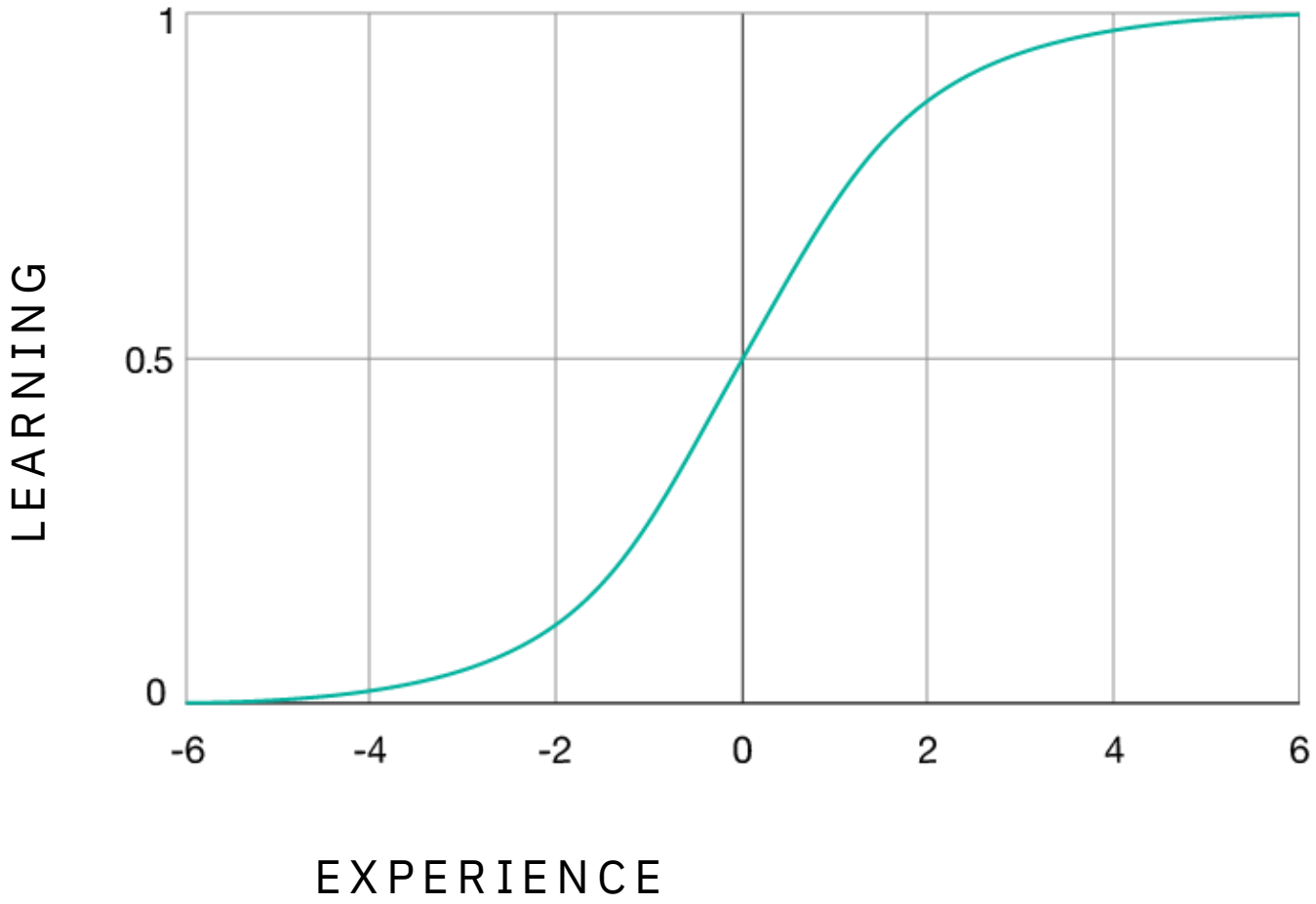
Linear Regression



Logistical Regression



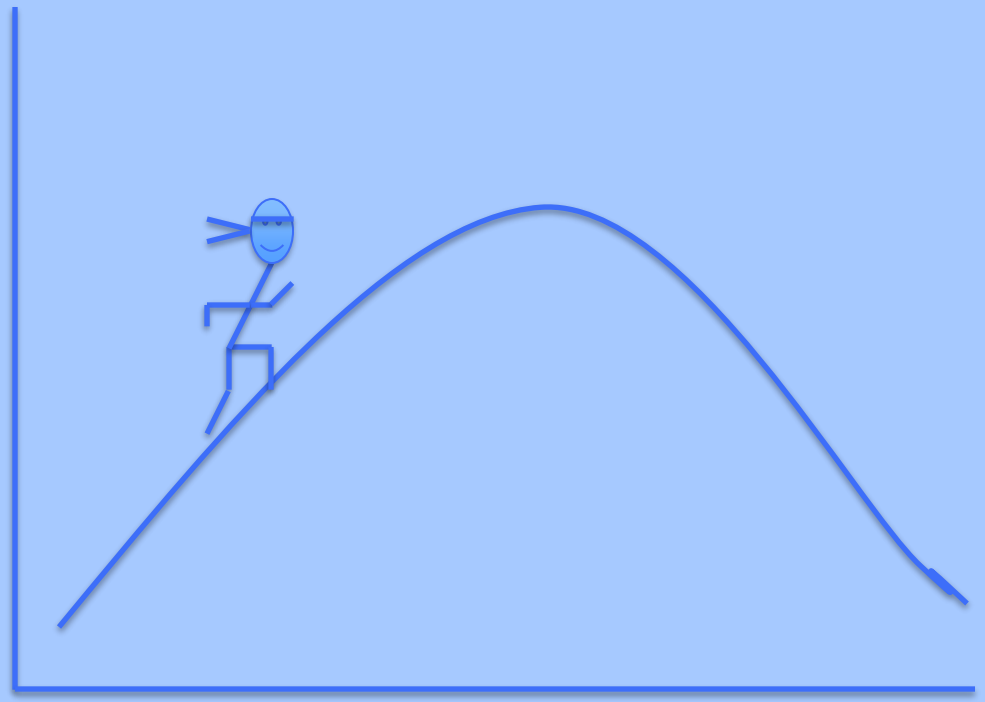
Everywhere you look there is an S-curve



$$f(x) = \frac{1}{1+e^{-x}}$$

The sigmoid function squashes a range from minus infinity to plus infinity to a **range 0 – 1**.

Machine learning is optimization



Climb to the top of the hill
Under two conditions:

a) Must do it blindfolded
b) In as minimum steps as possible

Predicting your grades using neural nets

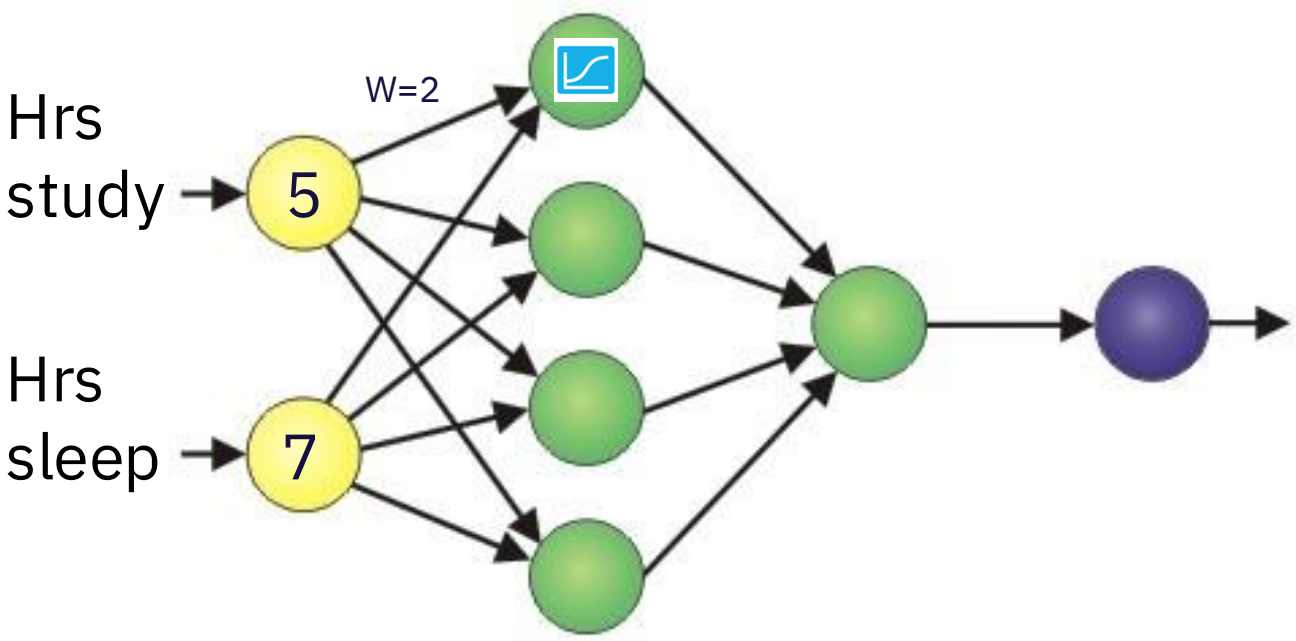
X	hrs study	hrs sleep	Grade	y
	7	5	78	
	6	8	93	
	8	2	67	
	5	5	?	$\hat{y}$

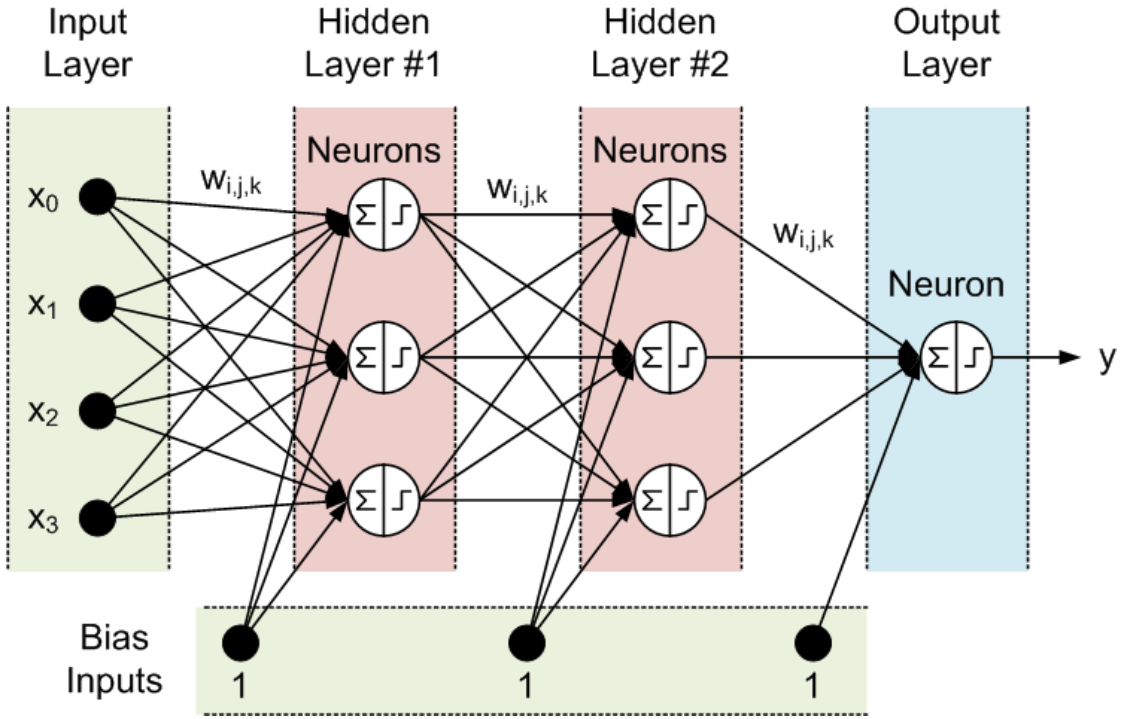
Supervised learning
This is a regression problem
Not a classification problem

$X_{norm} = x / \max(x)$

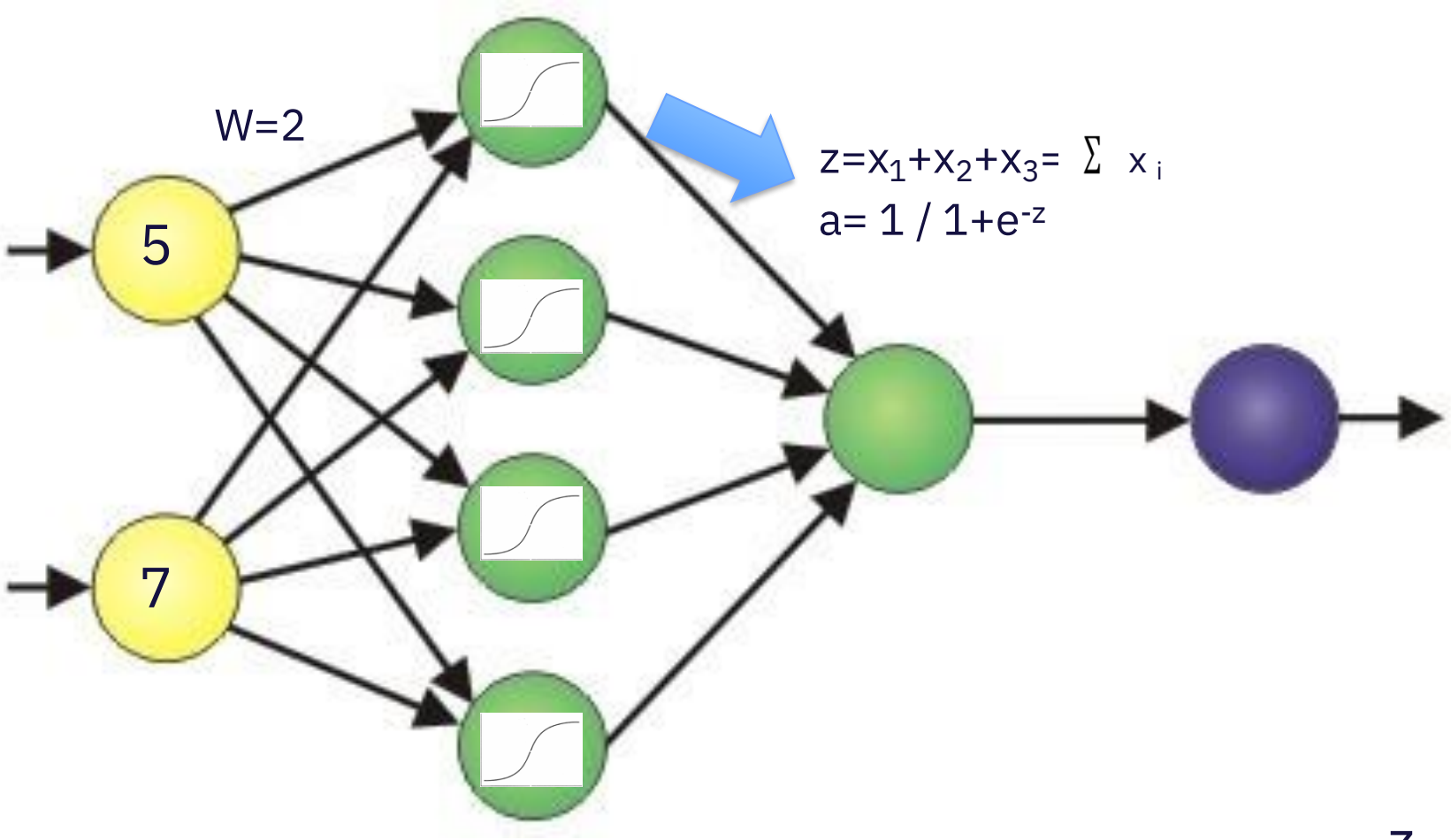
$Y_{norm} = y / \max(y)$

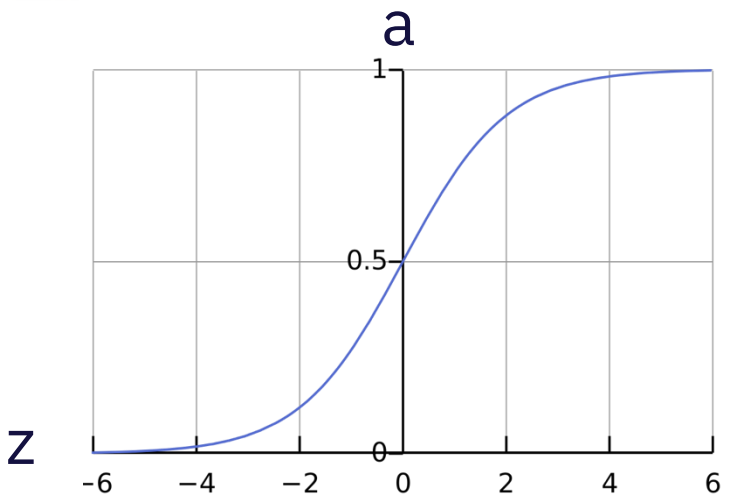
Basic structure of a neural net



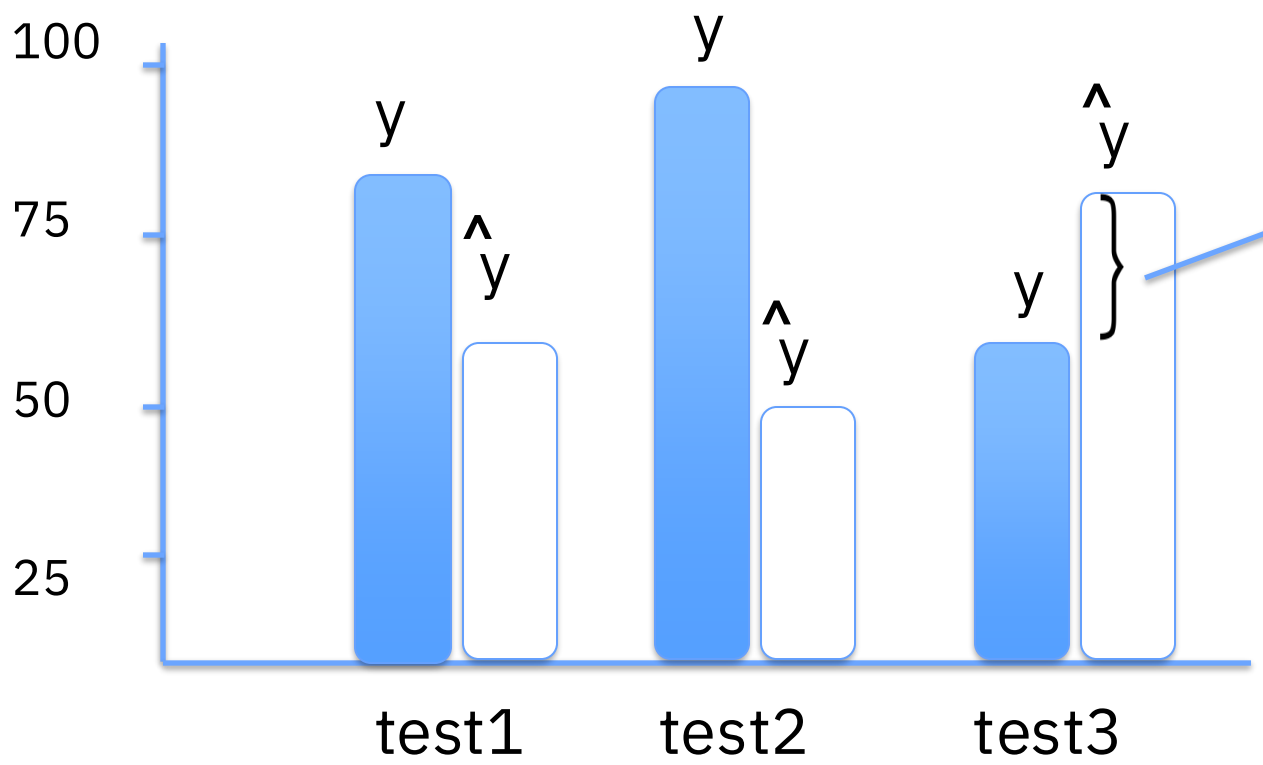


Sigmoid Operation





Training neural network = minimizing cost function

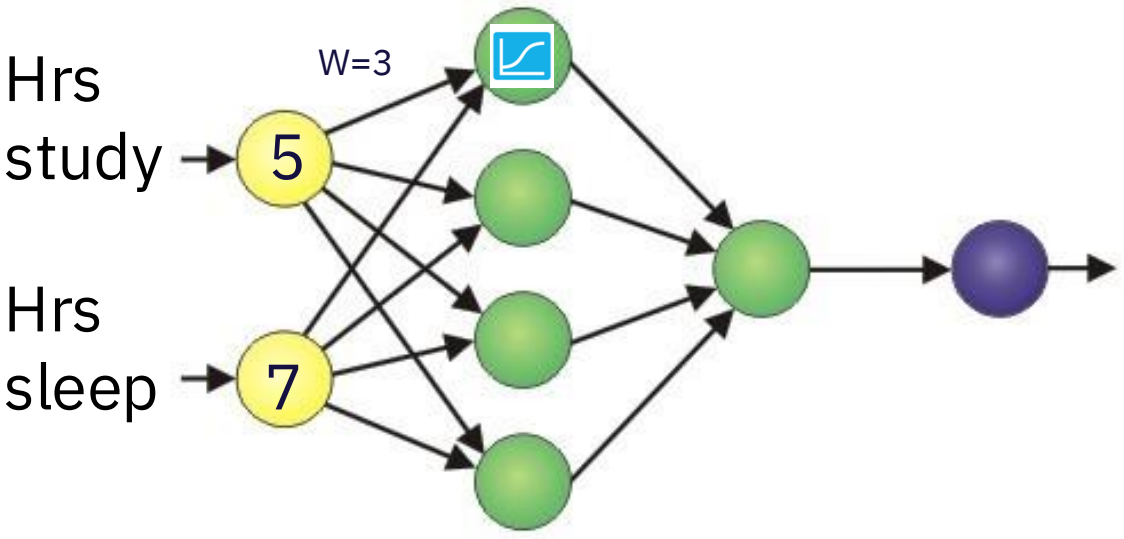


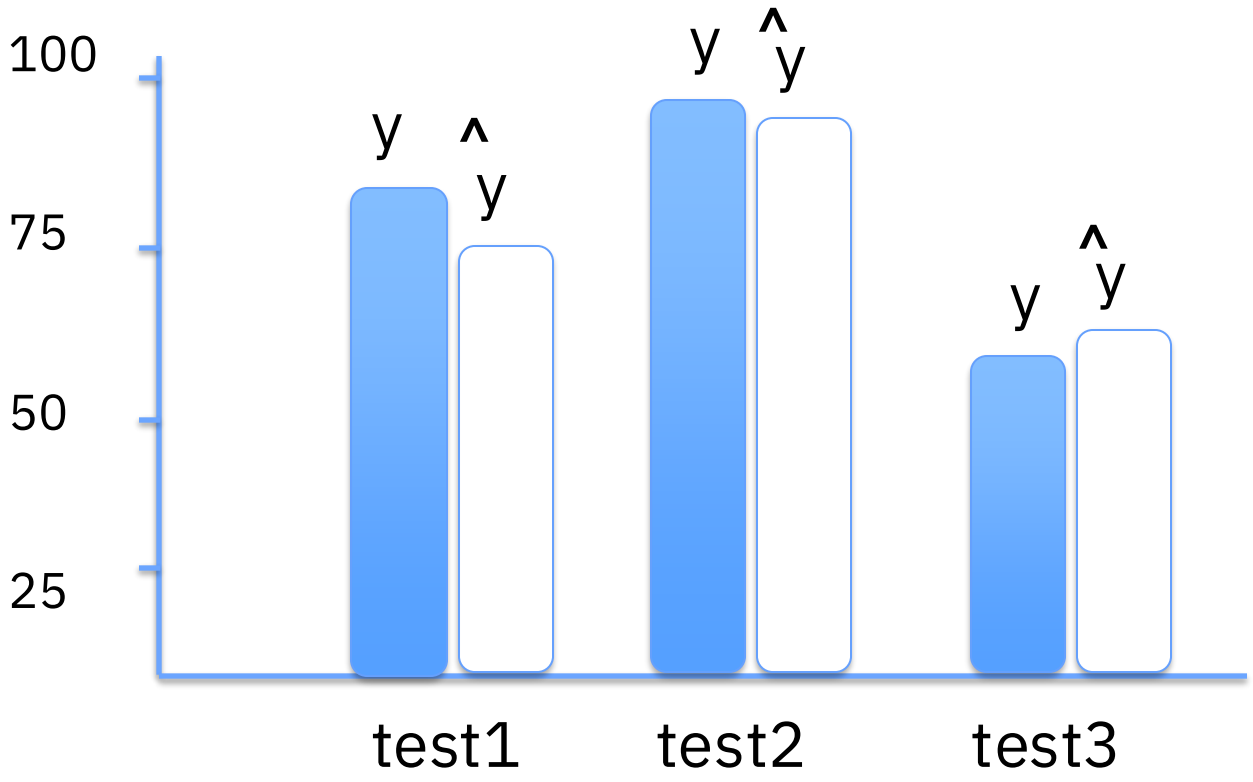
$$e = y - \hat{y}$$

$$j = cost = e_1^2 + e_2^2 + e_3^2$$

$$j = \sum \tfrac{1}{2} (y - y_{hat})^2$$

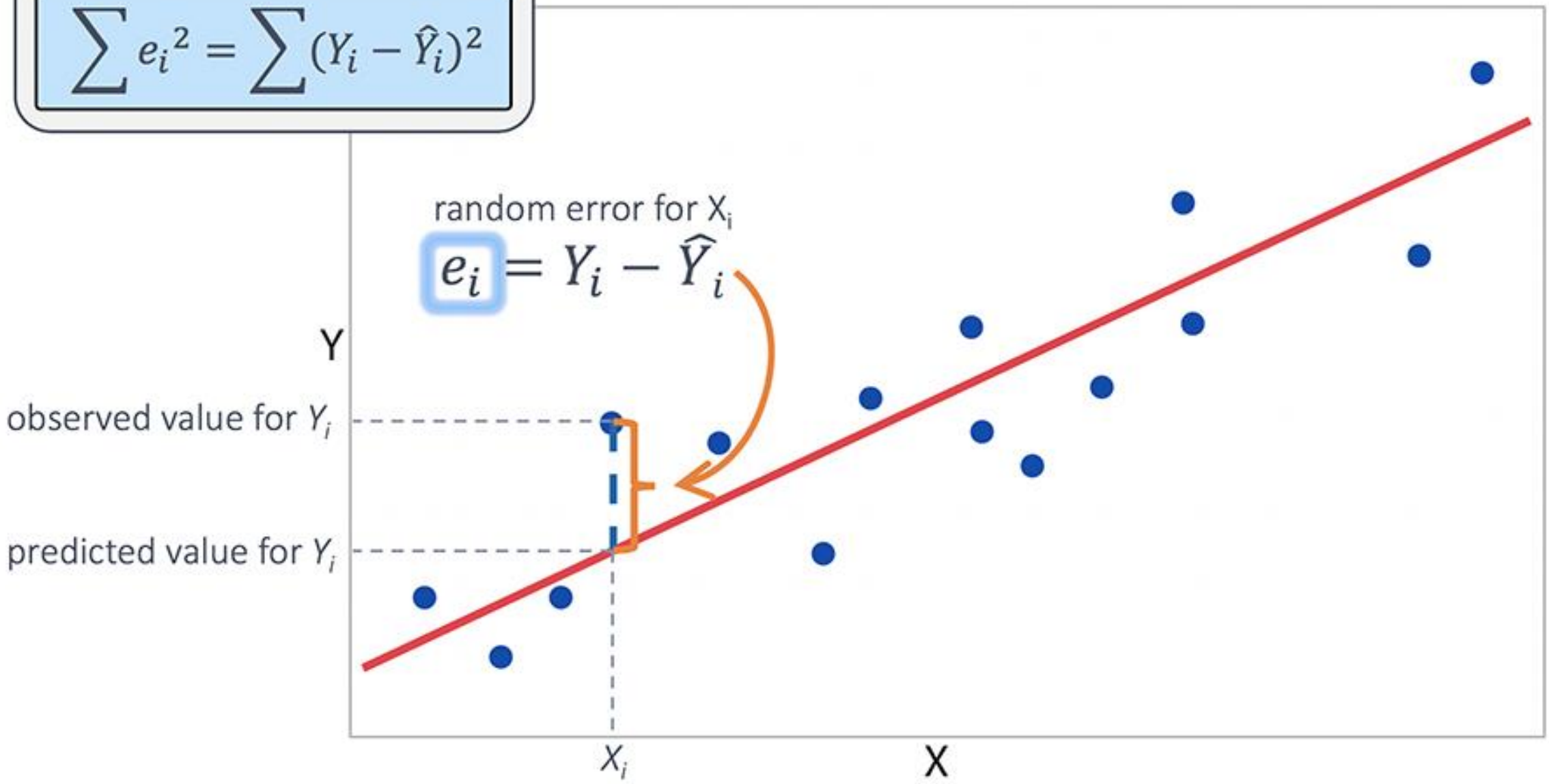
Backpropagation



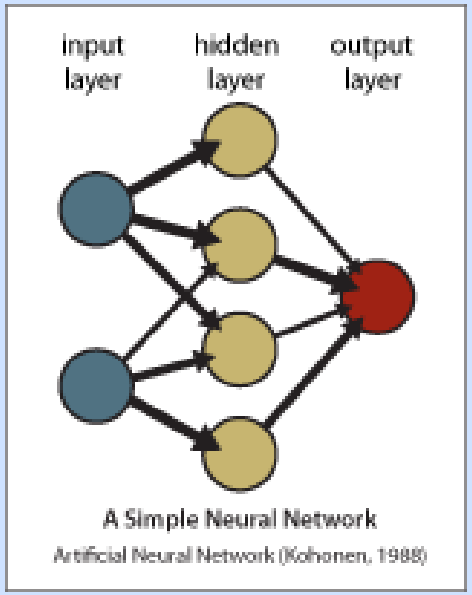


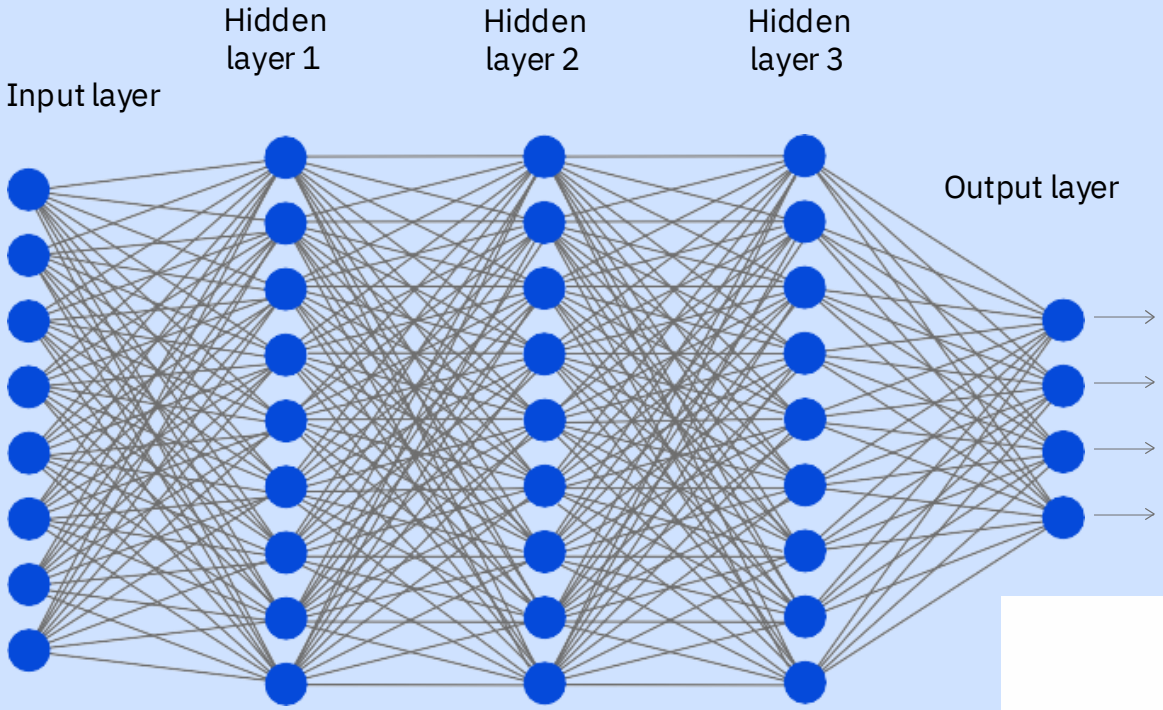
Method of Least Squares

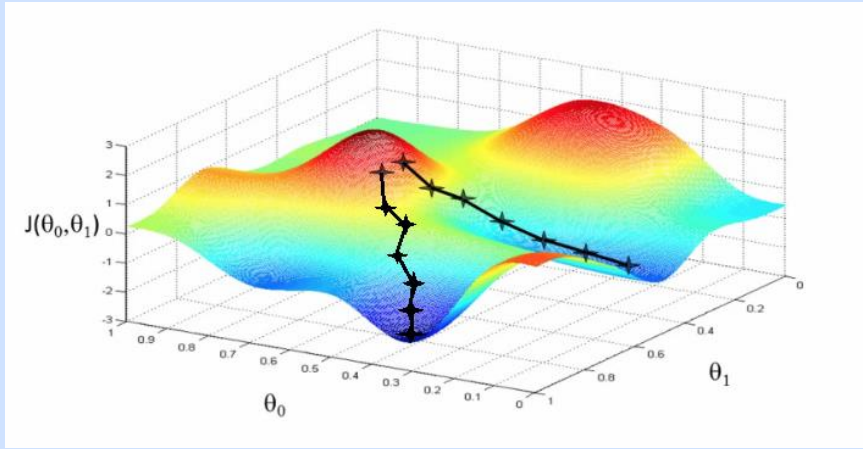
$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$



About neural networks





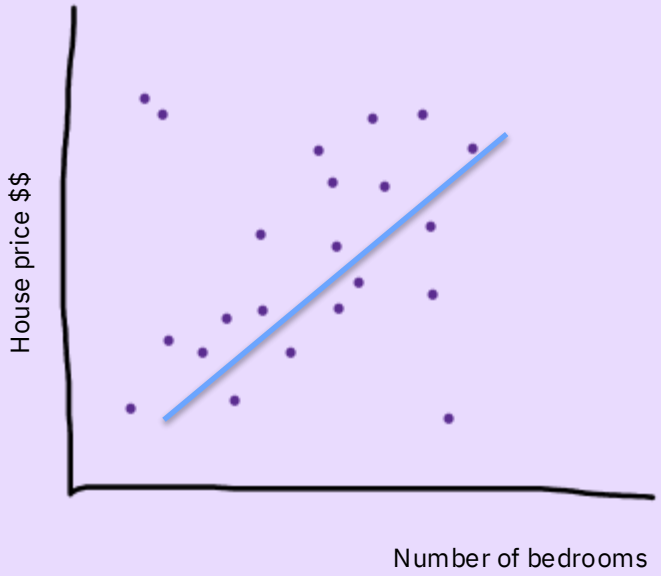


Dear students consider this scenario:

You are a Real Estate agent and your associate drew a **blue line** as the *prediction* line depicting How steeply (or not) the price of houses rise as the number of bedrooms increases.

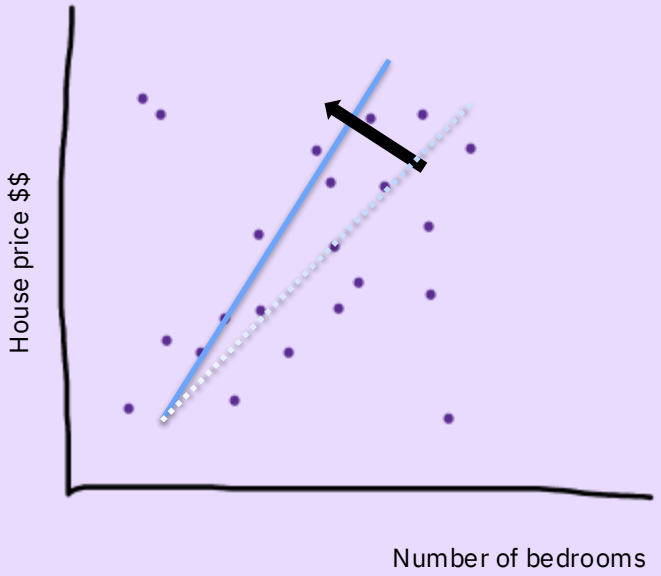
So, you ponder the prediction line which supposedly is playing fair with all those points: **The MSE calculation**.

And at a first glance, the slope should be steeper. →

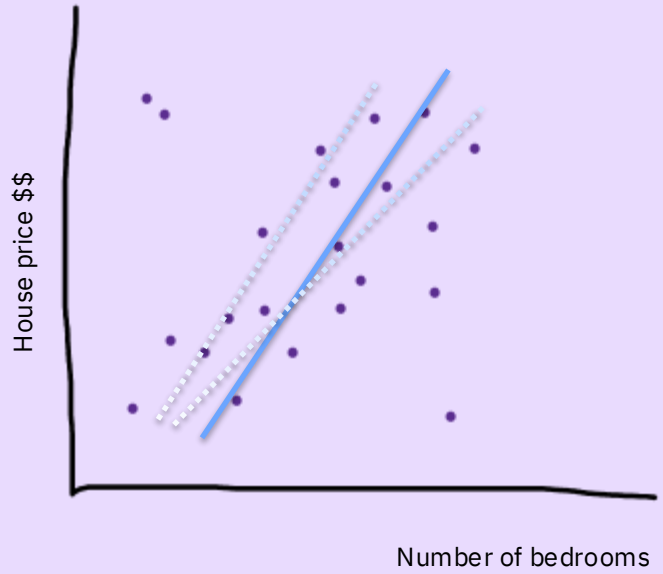


So, you went ahead and changed the slope. Slope means *gradient*, and gradient means ***learning rate.***

To make a long story short, when the slope changes, the **Weights** have something to do with it.



At a third glance, you decide not only to adjust the slope (weights) but also to move the entire curve around. This is when the **Bias** also comes into play to move the whole curve (or line).



Brute Statistics versus Artificial Neural Networks

Regression takes the data and tries to find the result that minimizes prediction mistakes, maximizing what is called goodness of fit.

A physicist, an engineer and a statistician go on a hunting trip....

Being precisely perfect on average can mean being actually wrong each time. Regression can keep missing several feet to the left or several feet to the right. Even if it averages out to be the correct answer, regression can mean never actually hitting the target.

Unlike regression, **machine learning** predictions might be wrong on average, but when the prediction miss, they often don't miss by much. Statisticians describe this as allowing some bias in exchange for reducing variance.

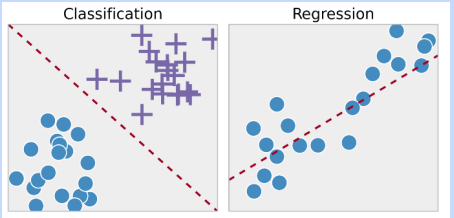
Inventing a new machine learning method involves proving that it works better in practice. In contrast, inventing a regression method requires first proving that it works in theory, it requires the articulation of a hypothesis.

Machine learning has less need to specify in advance what goes into the model and can accommodate the equivalent of much more complex models with many more interactions between variables.

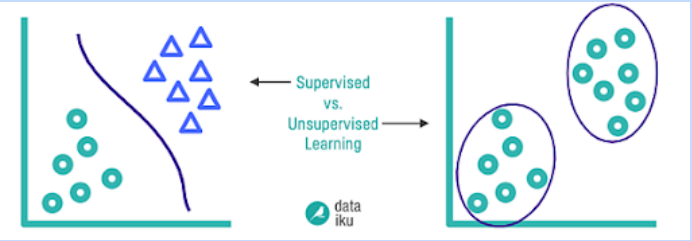
These excerpts are taken from the book **Prediction Machines** by Ajay Agrawal, Josh Gans and AVI Goldfarb

What are the three machine learning approaches?

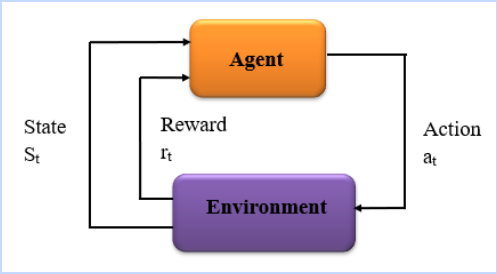
- Supervised learning



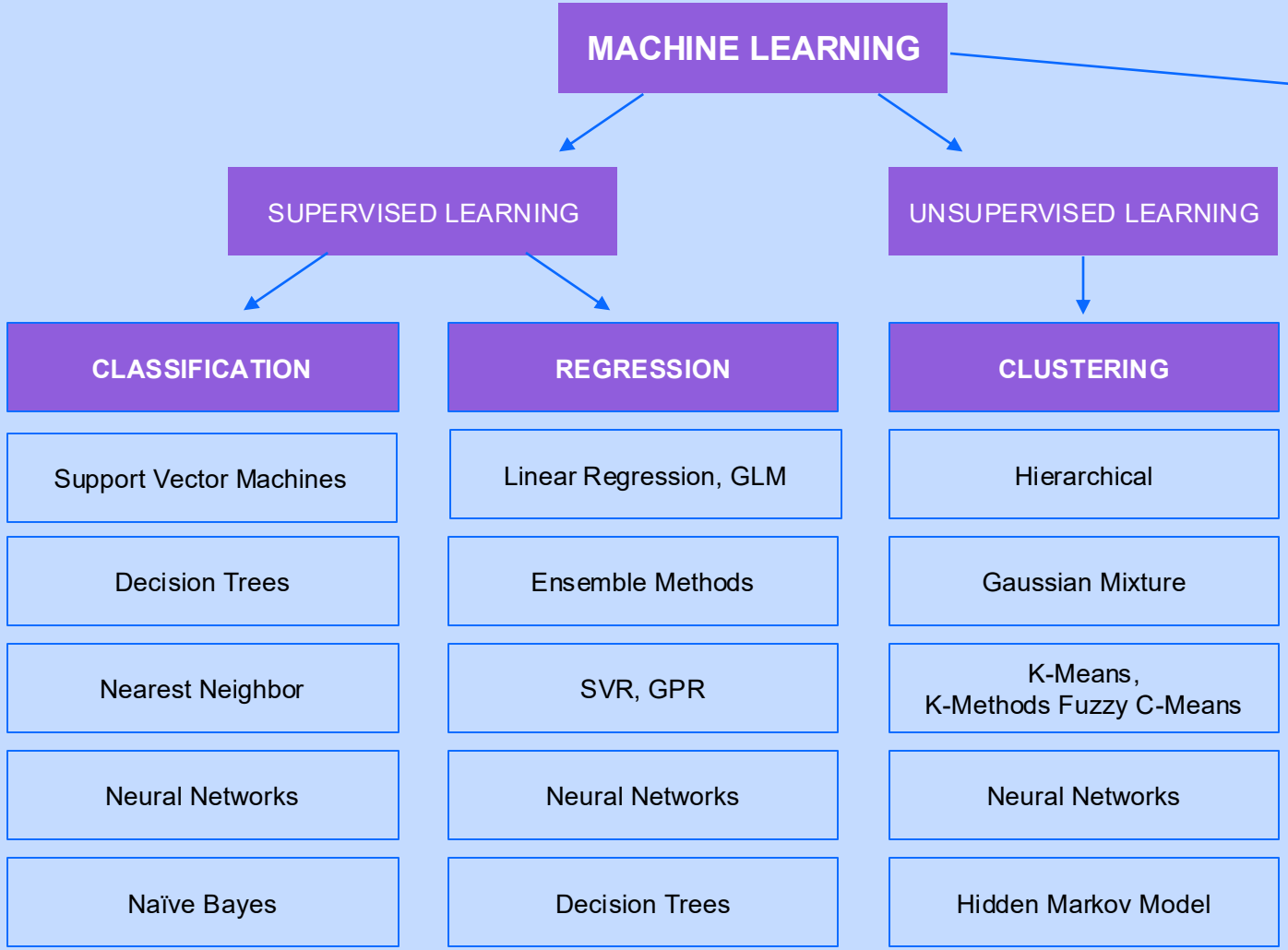
- Unsupervised learning

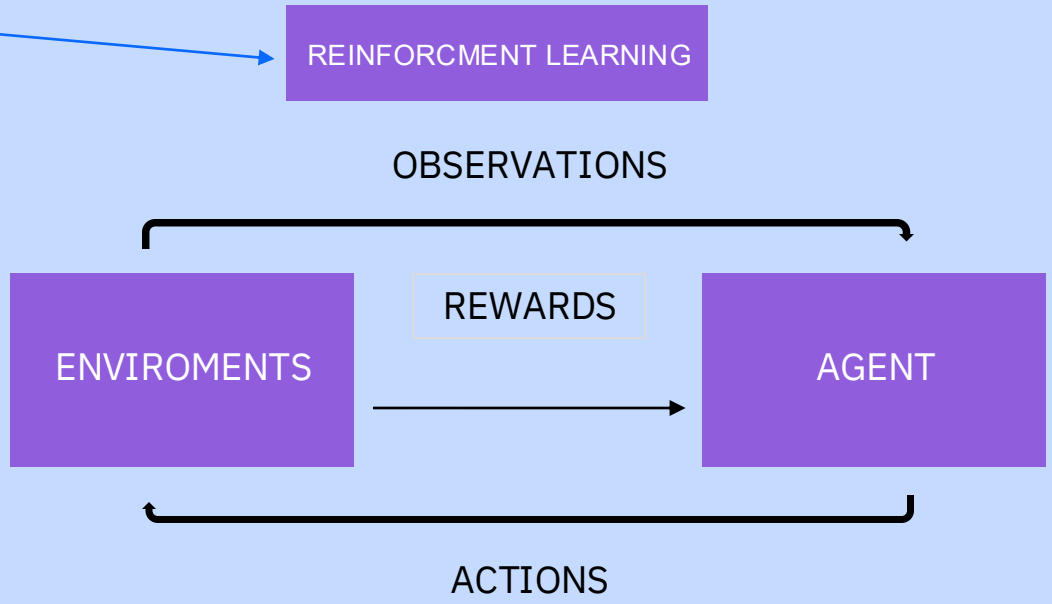


- Reinforcement learning

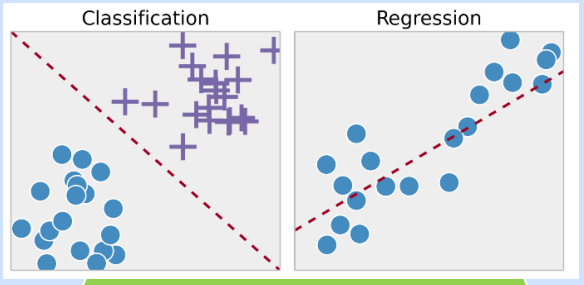


Algorithms used with machine learning methods



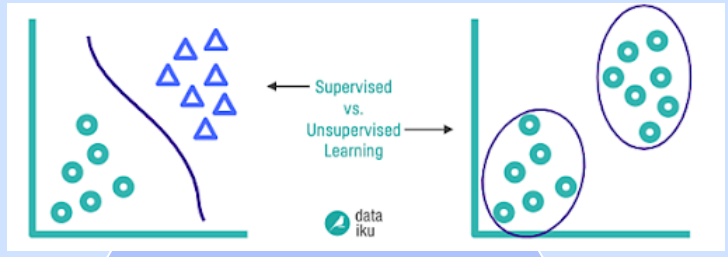


Machine Learning



Supervised

- Classification
  - Identity fraud detection
  - Image Classification
  - Customer retention
  - Diagnostics
- Regression
  - Weather forecasting
  - Market forecasting
  - Advertising popularity prediction
  - Estimating life expectancy



Unsupervised

- Dimensionality reduction
  - Big Data Visualization
  - Structure discovery
  - Meaningful compression
- Clustering
  - Recommender systems
  - Customer segmentation
  - Targeted Marketing

Supervised vs. Unsupervised Machine Learning

Parameters	Supervised ML techniques	Unsupervised ML techniques
Process	In a supervised learning model, input and output variables are given.	In unsupervised learning model, only input data will be given
Input data	Algorithms are trained using labeled data.	Algorithms are used against data which is not labeled
Algorithms used	Support vector machine, Neural network, Linear and logistics regression, random forest, and Classification trees.	Unsupervised algorithms can be divided into different categories: like Cluster algorithms, K-means, Hierarchical clustering, etc.
Computational complexity	Supervised learning is a simpler method.	Unsupervised learning is computationally complex
Use of data	Supervised learning model uses training data to learn a link between the input and the outputs.	Unsupervised learning does not use output data.
Accuracy of results	Highly accurate and trustworthy method.	Less accurate and trustworthy method.
Real-time learning	Learning method takes place offline.	Learning method takes place in real time.
Number of classes	Number of classes is known.	Number of classes is not known.
Main drawback	Classifying big data can be a real challenge in Supervised Learning.	You cannot get precise information regarding data sorting, and the output as data used in unsupervised learning is labeled and not known.

What is reinforcement learning

Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions.

For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.

In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike supervised learning.

Since there is no labeled data, so the agent is bound to learn by its experience only.

When do I use reinforcement learning

RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as **game-playing, robotics**, etc. The agent interacts with the environment and explores it by itself.

The primary goal of an agent in reinforcement learning is to improve the performance by getting the maximum positive rewards.

The agent learns with the process of hit and trial, and based on the experience, it learns to perform the task in a better way. Hence, we can say that ***"Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that."*** How a Robotic dog learns the movement of his arms is an example of Reinforcement learning.

The agent continues doing these three things (**take action, change state/remain in the same state, and get feedback**), and by doing these actions, the agent learns and explores the environment.

The agent learns that what actions lead to positive feedback or rewards and what actions lead to negative feedback penalty. As a positive reward, the agent gets a positive point, and as a penalty, it gets a negative point.

LECTURE 5

Machine Learning and Deep Learning

1. Machine Learning Explained
2. **Decision Tree Classifier**
3. Deep Learning Ecosystem
4. Resources

Predict if Nemra will commute to the office

Hard to guess under what conditions Nemra will work from home or drive to the office.

Let's divide and conquer:

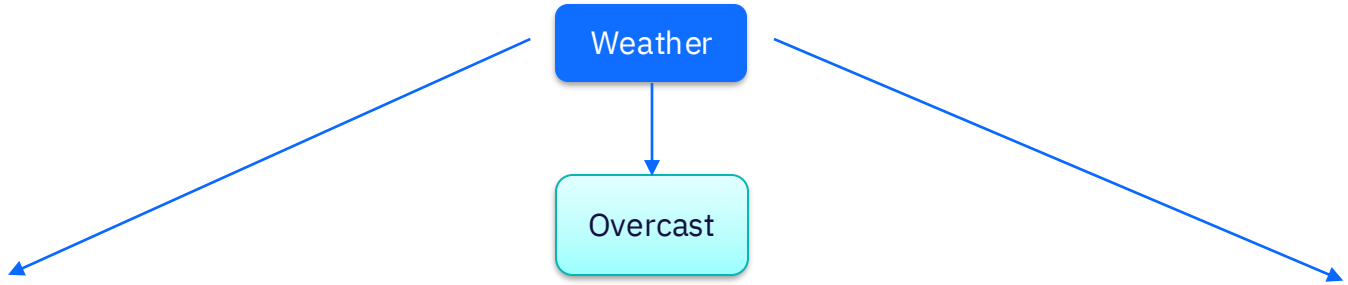
- Split into subsets
- Are the all pure (all yes or all no)
- If yes: stop
- If no: repeat

Day	Weather	Type of calls	Exercise	Commute?
D15	Rain	Video	Walk	**?**

Day	Weather	Type of calls	Exercise	Commutes
D1	Sunny	Video Conference	Walk	No
D2	Sunny	Video Conference	Gym	No
D3	Overcast	Video Conference	Walk	Yes
D4	Rain	Video Conference	Walk	Yes
D5	Rain	Telephone	Walk	Yes
D6	Rain	Telephone	Gym	No
D7	Overcast	Telephone	Gym	Yes
D8	Sunny	Video Conference	Walk	No
D9	Sunny	Telephone	Walk	Yes
D10	Rain	Telephone	Walk	Yes
D11	Sunny	Telephone	Gym	Yes
D12	Overcast	Video Conference	Gym	Yes
D13	Overcast	Telephone	Walk	Yes
D14	Rain	Video Conference	Gym	No

The Decision Tree

9 yes / 5 no



Day	Weather	Type of calls	Exercise
D1	Sunny	Video	Walk
D2	Sunny	Video	Gym
D8	Sunny	Video	Walk
D9	Sunny	Telephone	Walk
D11	Sunny	Telephone	Gym

2 yes / 3 no
Split further

Day	Weather	Type of calls	Exercise
D3	Overcast	Video	Walk
D7	Overcast	Telephone	Gym
D12	Overcast	Video	Gym
D13	Overcast	Telephone	Walk

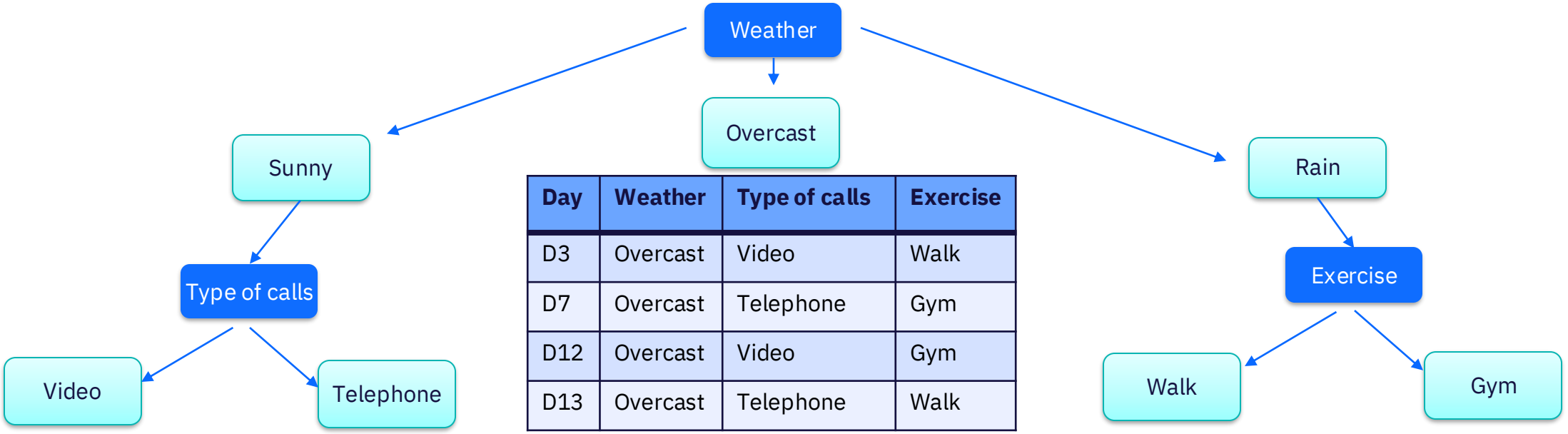
4 yes / 0 no
Pure subset

Day	Weather	Type of calls	Exercise
D4	Rain	Video	Walk
D5	Rain	Telephone	Walk
D6	Rain	Telephone	Gym
D10	Rain	Telephone	Walk
D14	Rain	Video	Gym

3 yes / 2 no
Split further

If Entropy, then branch further

9 yes / 5 no



Day	Weather	Type of calls	Exercise
D3	Overcast	Video	Walk
D7	Overcast	Telephone	Gym
D12	Overcast	Video	Gym
D13	Overcast	Telephone	Walk

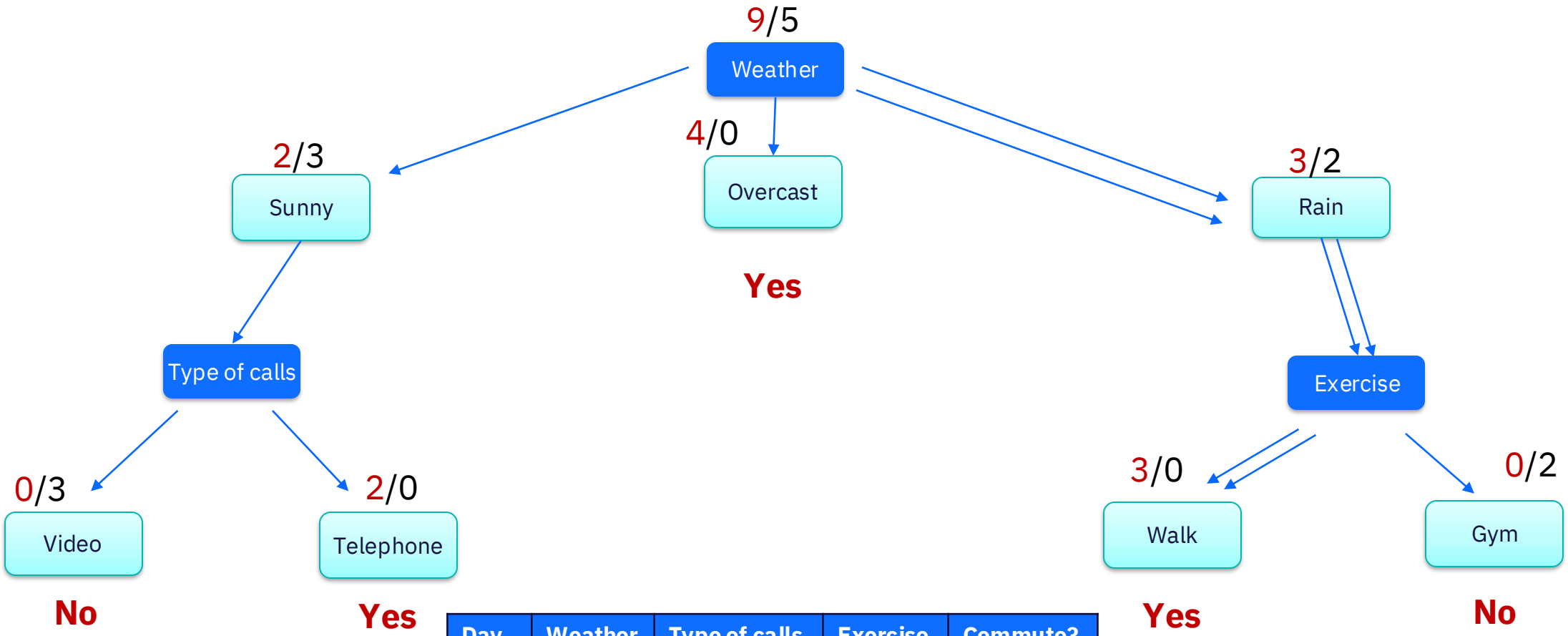
Day	Type of calls	Exercise
D1	Video	Walk
D2	Video	Gym
D8	Video	Walk

Day	Type of calls	Exercise
D9	Telephone	Walk
D11	Telephone	Gym

Day	Type of calls	Exercise
D4	Video	Walk
D5	Telephone	Walk
D10	Telephone	Walk

Day	Type of calls	Exercise
D6	Telephone	Gym
D14	Video	Gym

The Prediction



Day	Weather	Type of calls	Exercise	Commute?
D15	Rain	Video	Walk	? **Yes**

New data: Day 15

LECTURE 3

Machine Learning and Deep Learning

1. Machine Learning Explained
2. Decision Tree Classifier
3. Deep Learning Explained
4. **Deep Learning Ecosystem**
5. Summary & Resources

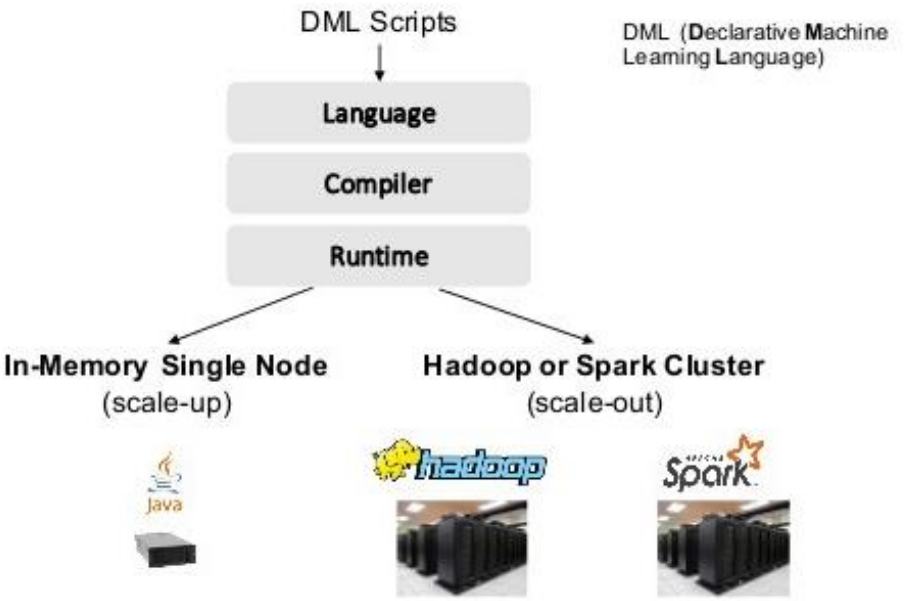
Deep learning ecosystem

Platform as a Service Providers:
Deep learning services included as part of PaaS solutions. Technologies like IBM Cloud, Microsoft Azure, Amazon AWS or Google Developer Cloud.

Deep Learning Frameworks:
Libraries and programming models that enable the fundamental constructs to build deep learning applications.

Deep learning framework: Apache SystemML (ML)

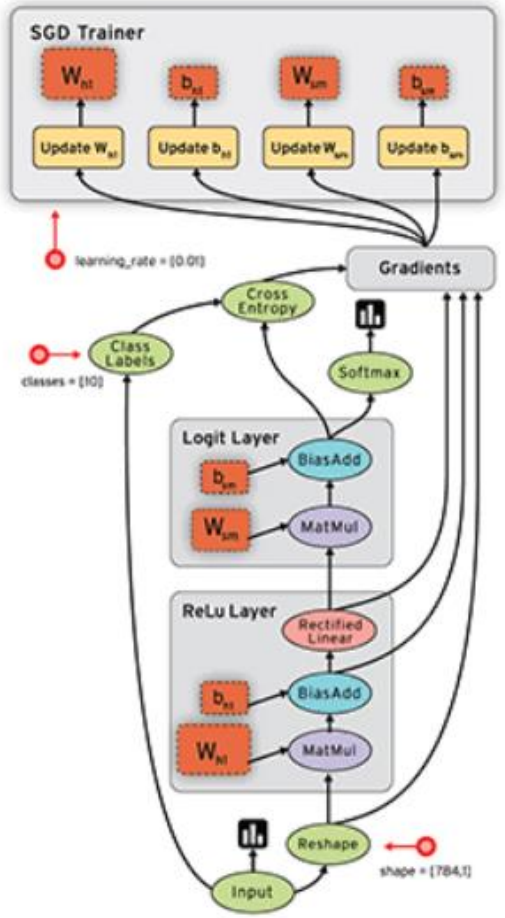
High-Level SystemML Architecture



The Apache SystemML language, Declarative Machine Learning (DML), includes linear algebra primitives, statistical functions, and ML-specific constructs that make it easier and more natural to express ML algorithms.

DML significantly increases the productivity of data scientists by providing full flexibility in expressing custom analytics as well as data independence from the underlying input formats and physical data representations.

Deep learning framework: TensorFlow



Google's TensorFlow deep learning framework was developed originally by the Google Brain Team for conducting research in machine learning and deep neural networks.

The framework's name is derived from the fact that it uses data flow graphs, where nodes represent a computation and edges represent the flow of information—in Tensor form—from one node to another.

Deep learning framework: Torch

Torch was based upon the scripting language Lua, which was designed to be portable, fast, extensible, and easy to use with an easy-to-use syntax.

Torch features a large number of community-contributed packages, giving Torch a versatile range of support and functionality.

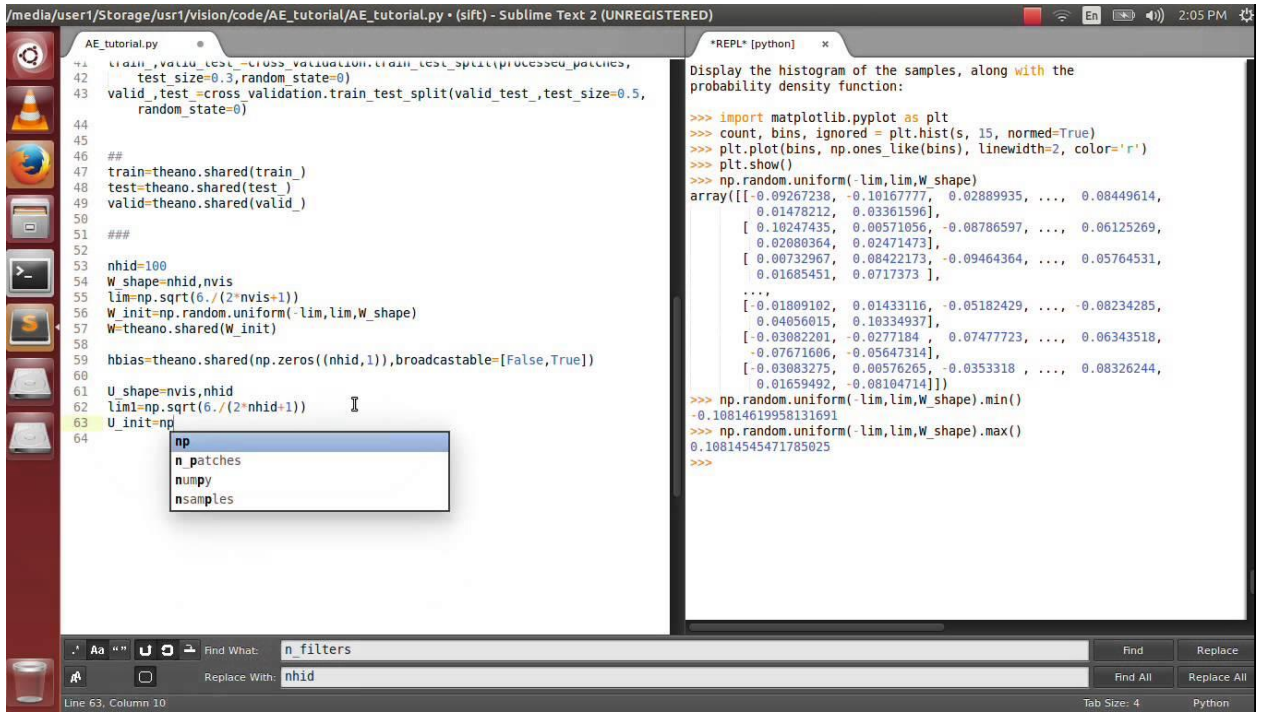
Simple NN in Torch

```
-- create closure to evaluate f(X) -- and df/dX
local feval = function(x)
   -- get new parameters
   if x ~= parameters then
      parameters:copy(x)
   end

   -- reset gradients
   gradParameters:zero()

   -- f is the average of all criterions
   local f = 0
```

Deep learning framework: Theano



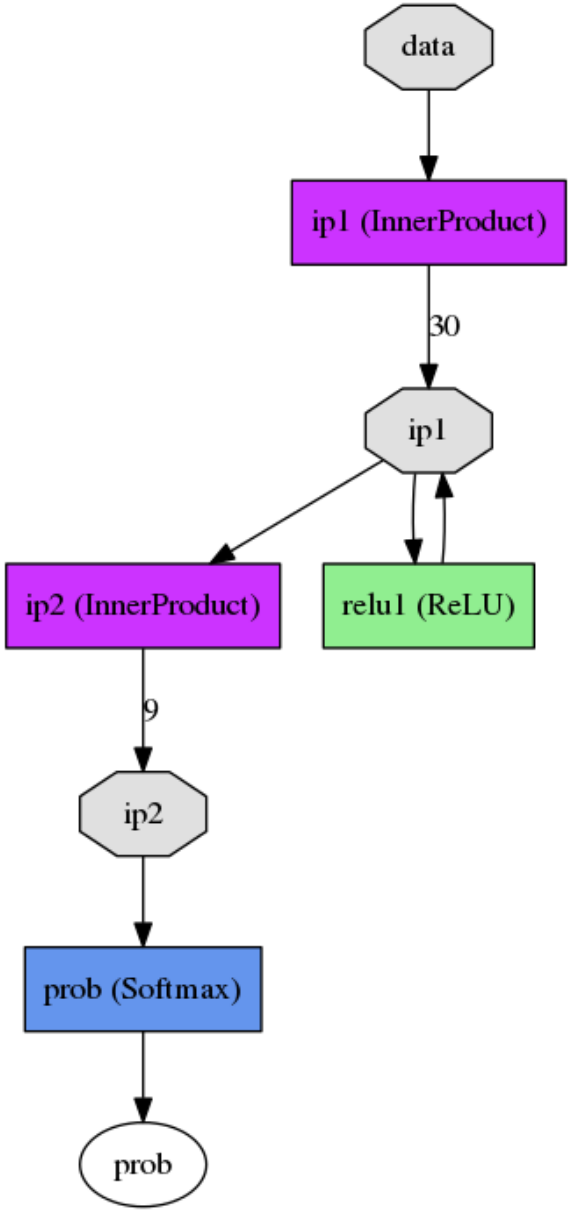
Very popular within the academic research community, Theano is considered grand-daddy of deep-learning frameworks, which is written in Python.

Theano is a library that handles multidimensional arrays, like Numpy. Used with other libs, it is well suited to data exploration and intended for research.

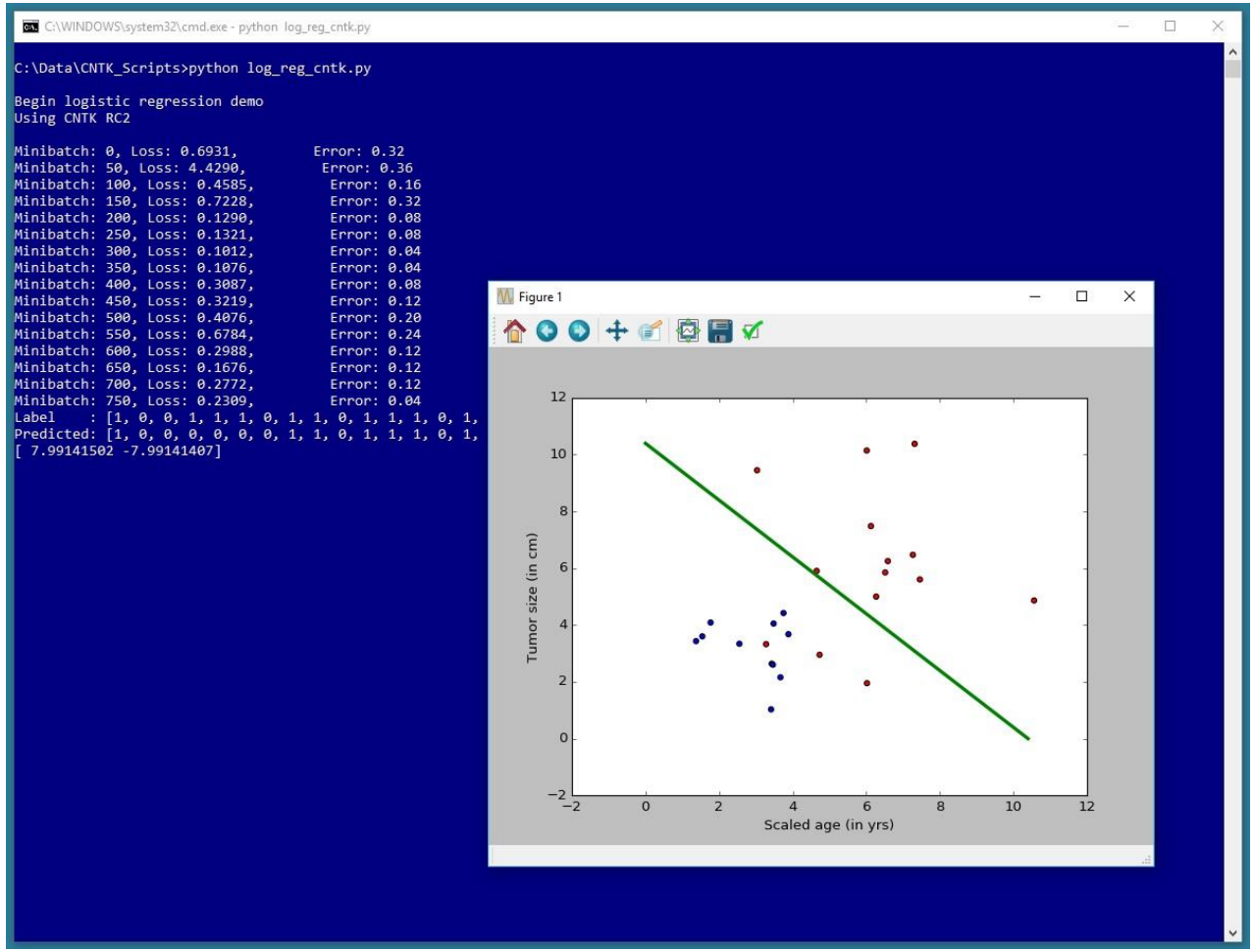
Deep learning framework: Caffe

Caffe is a well-known and widely used machine-vision library that ported Matlab's implementation of fast convolutional nets to C and C++. Caffe was developed at the Berkeley Vision and Learning Center (BVLC).

Caffe is useful for performing image analysis (Convolutional Neural Networks, or CNNs) and regional analysis within images using convolutional neural networks (Regions with Convolutional Neural Networks, or RCNNs).



Deep learning framework: CNTK



CNTK is Microsoft's open-source deep-learning framework. The acronym stands for "Computational Network Toolkit."

While CNTK appears to have a permissive license, it has not adopted one of the more conventional licenses, such as ASF 2.0, BSD or MIT.

LECTURE 3

Machine Learning and Deep Learning

1. Machine Learning Explained
2. Decision Tree Classifier
3. Deep Learning Ecosystem
4. **Resources**

Resources *page 1 of 2*

1. Machine Learning, Deep Learning 101
   https://developer.ibm.com/articles/l-machine-learning-deep-learning-trs/?mhq=deep%20learning%20101&mhsrc=ibmsearch_a

2. Microsoft Softbank Build Cloud Robotics Retail Business
   https://winbuzzer.com/2016/03/10/microsoft-softbank-build-cloud-robotics-retail-business-xcxwbn/

3. Curiosity Photos Show Martian Dust Storm Growing
   https://www.nasa.gov/feature/goddard/2018/curiosity-photos-show-martian-dust-storm-growing

4. IoT Deep Learning Anomaly Detection
   https://developer.ibm.com/tutorials/iot-deep-learning-anomaly-detection-1/

5. Backpropagation
   https://en.wikipedia.org/wiki/Backpropagation

6. Connectionism
   https://en.wikipedia.org/wiki/Connectionism

7. Gradient Descent
   https://en.wikipedia.org/wiki/Gradient_descent

Resources *page 2 of 2*

1. Neural Networks with Scikit-Learn
   https://stackabuse.com/introduction-to-neural-networks-with-scikit-learn/

2. Apache
   https://systemml.apache.org/

3. TensorFlow
   https://developer.ibm.com/articles/cc-get-started-tensorflow/

4. Torch
   http://torch.ch/

5. Theano
   http://deeplearning.net/software/theano/

6. Caffe
   http://caffe.berkeleyvision.org/

7. CNTK
   https://github.com/Microsoft/CNTK