Artificial Intelligence

Supervised Learning

Following Slides contain material from Bing Liu, Mackassy, Stuart Russell, Berrin Yanikoglu

Example: Admit in ICU

- An emergency room in a hospital measures 17 variables (e.g., blood pressure, age, etc) of newly admitted patients.
- A decision is needed: whether to put a new patient in an intensive-care unit.
- Due to the high cost of ICU, those patients who may survive less than a month are given higher priority.
- Problem: to predict high-risk patients and discriminate them from low-risk patients. Data???
- Images (object identification), textual (chats, text generation, word prediction, sentence completion, audio, speech recognize, speaker identify, emotion identification), EE brain signals, medical, integers,
- Tabular data (finance bank data, loan, id, demographic (age, gender, ethnicity, education, classes, attendance, subs, GPA, assign marks, online learning (clickstream, edx, udemy, log data, clicks,)))

Example: Loan Approval

- A credit card company receives thousands of applications for new cards. Each application contains information about an applicant,
  - age
  - Marital status
  - annual salary
  - outstanding debts
  - credit rating
  - etc.
- Problem: to decide whether an application should be approved, or to classify applications into two categories, approved and not approved.

Example: Spam Filter

Input: email

Output: spam/ham

Setup:

- Get a large collection of example emails, each labeled "spam" or "ham"
- Note: someone has to hand label all this data!
- Want to learn to predict labels of new, future emails

Features: The attributes used to make the ham / spam decision

- Words: FREE!
- Text Patterns: $dd, CAPS
- Non-text: SenderInContacts
- …

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Example: Digit Recognition

Input: images / pixel grids

Output: a digit 0-9

Setup:

- Get a large collection of example images, each labeled with a digit
- Note: someone has to hand label all this data!
- Want to learn to predict labels of new, future digit images

Features: The attributes used to make the digit decision

- Pixels: (6,8)=ON
- Shape Patterns: NumComponents, AspectRatio, NumLoops
- …

0

1

2

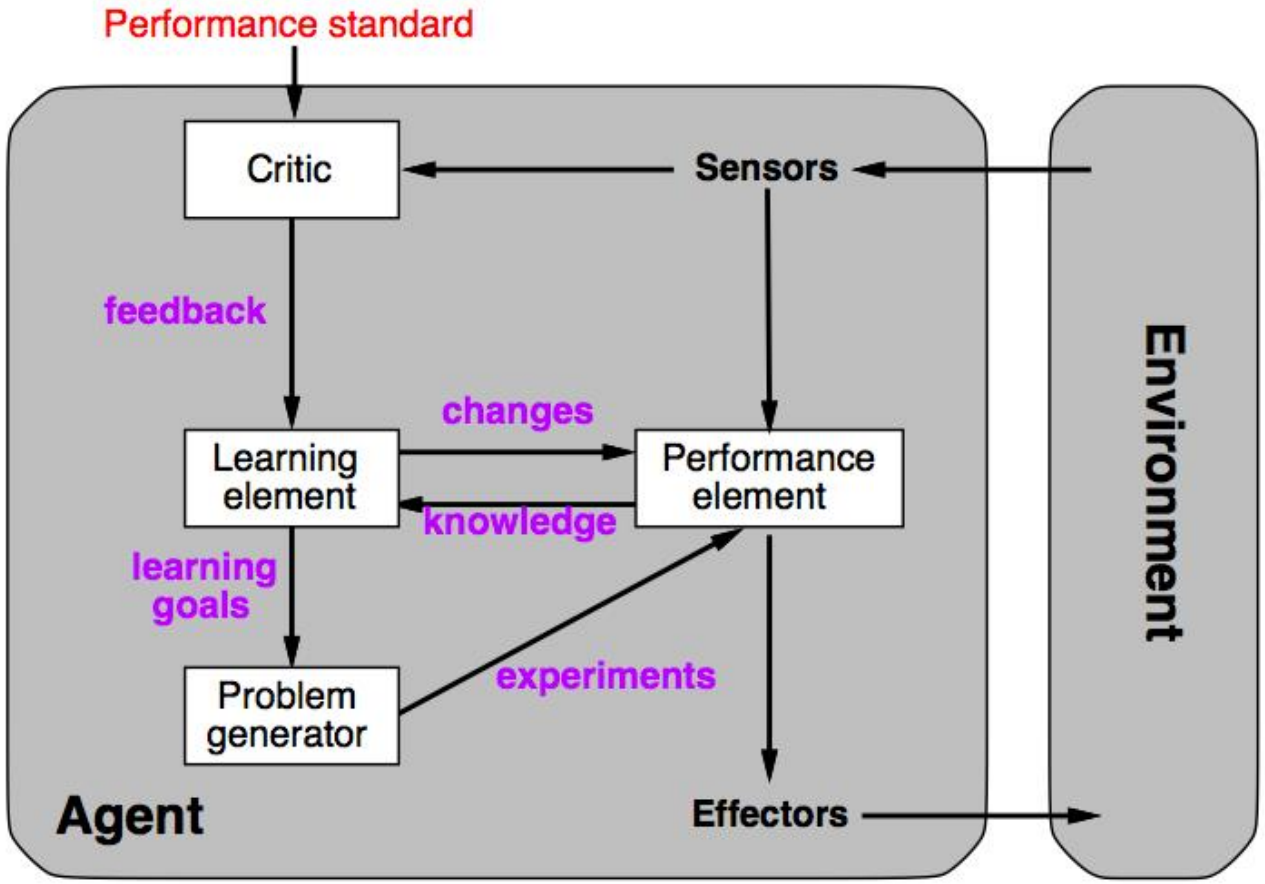
1

??

Example: Surveillance Videos

Example: Pose Transfer

Structure of Learning Agent

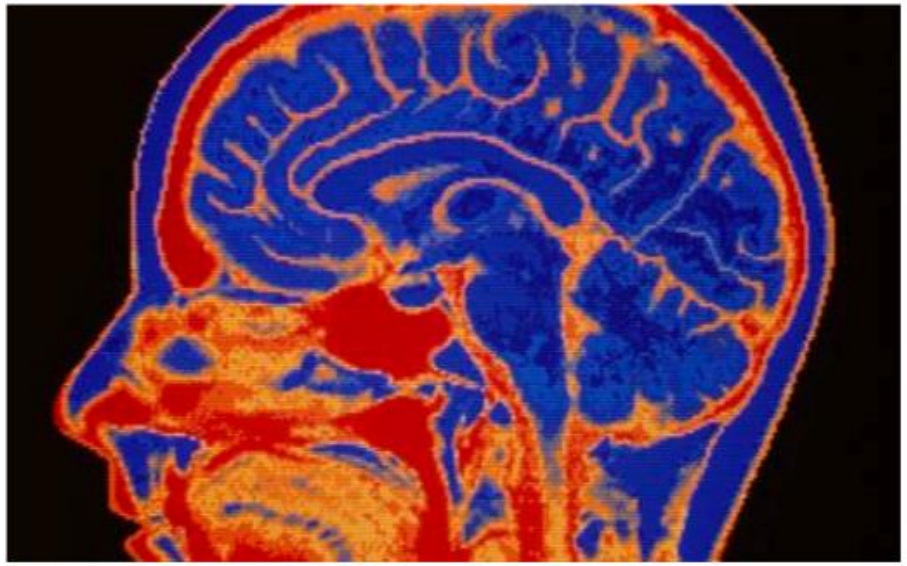


Machine Learning

The Telegraph

New IBM microchip to 'rival' the human brain

IBM has developed silicon microchips that it says mimic how the brain works and "rewire" themselves in response to new information.



A cross-section of the human brain

By Christopher Williams (http://www.telegraph.co.uk/journalists/christopher-williams/) , Technology Correspondent

7:00AM BST 19 Aug 2011

Machine learning and our focus

- Like human learning from past experiences.
- A computer does not have “experiences”.
- A computer system learns from data, which represent some “past experiences” of an application domain.
- Our focus: learn a target function that can be used to predict the values of a discrete class attribute, e.g., approve or not-approved, high-risk or low risk, spam or not spam, and digit or not digit.
- The task is commonly called: Supervised learning or classification

The data and the goal

- Data: A set of data records (also called examples, instances or cases) described by
  - *k* attributes: A_1, A_2, … A_k.
  - a class: Each example is labelled with a pre-defined class.
- Goal: To learn a classification model from the data that can be used to predict the classes of new (future, or test) cases/instances.

An example: data (loan application)

Approved or not

ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	**No**
2	young	false	false	good	**No**
3	young	true	false	good	**Yes**
4	young	true	true	fair	**Yes**
5	young	false	false	fair	**No**
6	middle	false	false	fair	**No**
7	middle	false	false	good	**No**
8	middle	true	true	good	**Yes**
9	middle	false	true	excellent	**Yes**
10	middle	false	true	excellent	**Yes**
11	old	false	true	excellent	**Yes**
12	old	false	true	good	**Yes**
13	old	true	false	good	**Yes**
14	old	true	false	excellent	**Yes**
15	old	false	false	fair	**No**

An example: the learning task

- Learn a classification model from the data
- Use the model to classify future loan applications into
  - Yes (approved) and
  - No (not approved)
- What is the class for following case/instance?

Age	Has_Job	Own_house	Credit-Rating	Class
young	false	false	good	?

Supervised vs Unsupervised Learning

- **Supervised Learning:** Data and corresponding labels are given
- **Unsupervised Learning:** Only data is given, no labels provided
- **Semi-supervised Learning:** Some (if not all) labels are present
- **Reinforcement Learning**: An agent interacting with the world makes observations, takes actions, and is rewarded or punished; it should learn to choose actions in such a way as to obtain a lot of reward

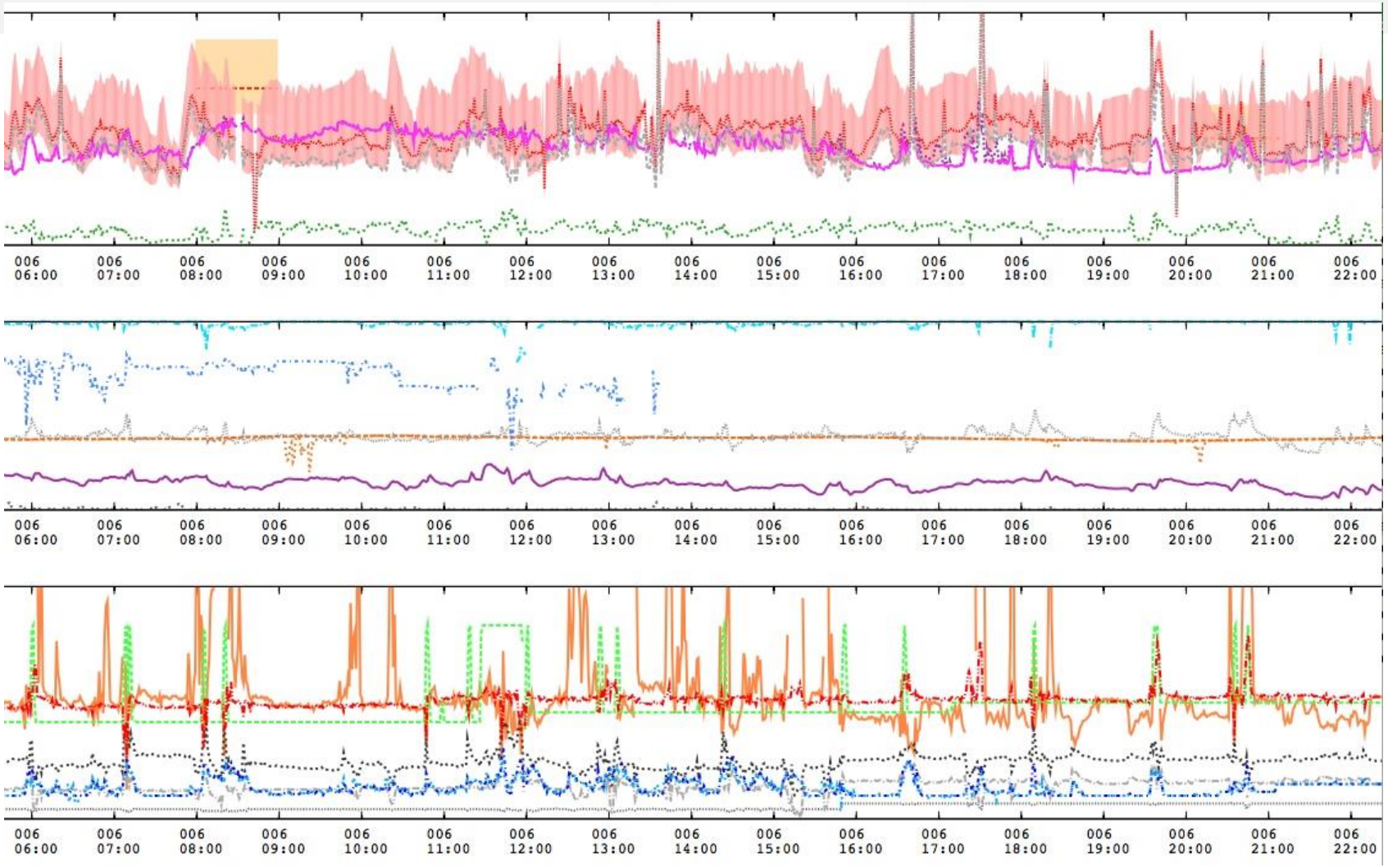
Lots of data

- Web: estimated Google index 45 billion pages
- Clickstream data: 10-100 TB/day
- Transaction data: 5-50 TB/day
- Satellite image feeds: ~1TB/day/satellite
- Sensor networks/arrays
  - CERN Large Hadron Collider ~100 petabytes/day
- Biological data: 1-10TB/day/sequencer
- TV: 2TB/day/channel; YouTube 4TB/day uploaded
- Digitized telephony: ~100 petabytes/day

ICU

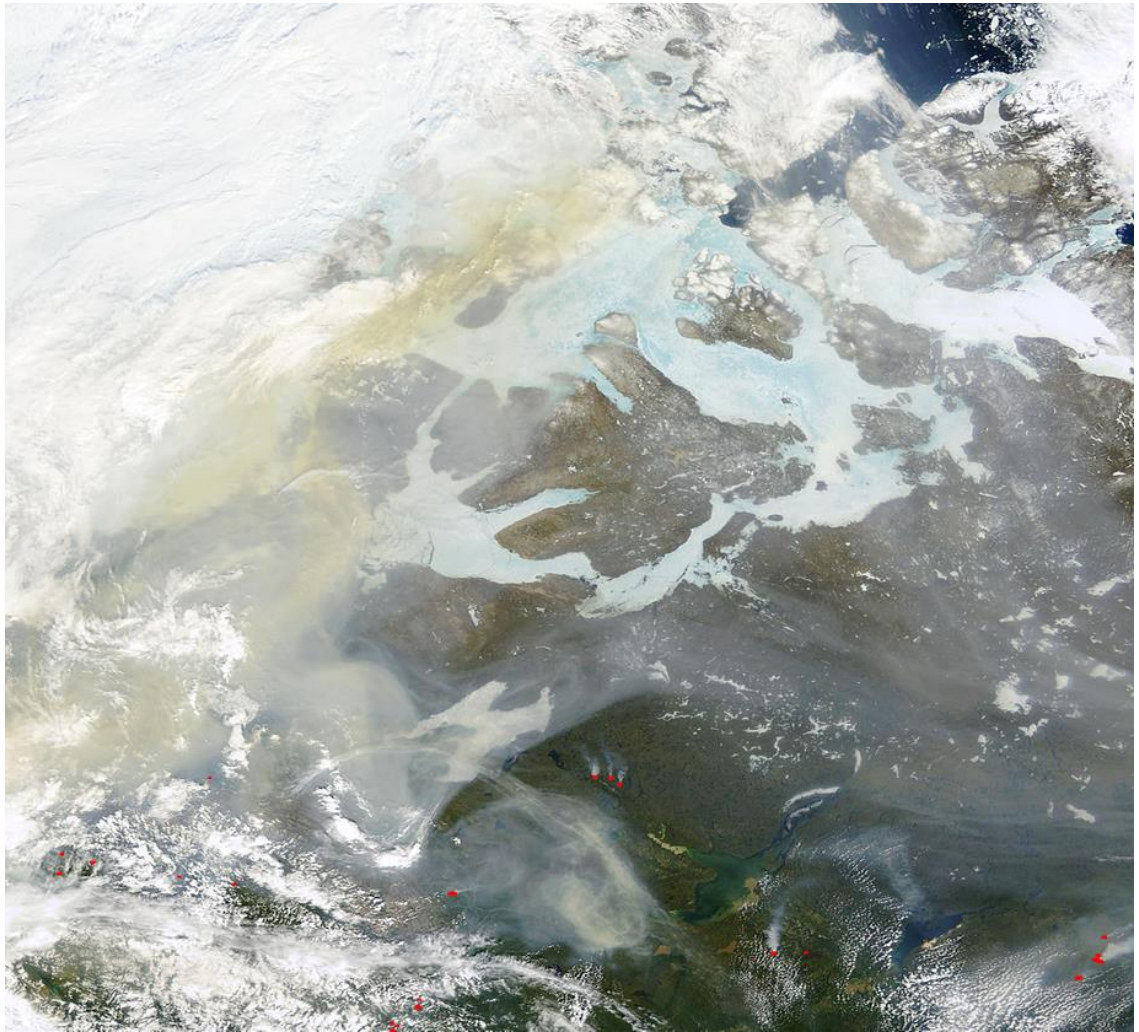
This is what an ICU looks like:
ventilator, fluids, monitors;
~200 medical procedures per day, many potentially fatal

Real data are messy



Arterial blood pressure (high/low/mean) 1s

Application: satellite image analysis



Application: Discovering DNA motifs

...TTGGAACAACCATGCACGGTTGATTCGTGCCTGTGACCGCGCGCCTCACACGGAAGACGCAGCCACCGGTTGTGATG

TCATAGGGAATTCCCCATGTCGTGAATAATGCCTCGAATGATGAGTAATAGTAAAACGCAGGGGAGGTTCTTCAGTAGTA

TCAATATGAGACACATACAAACGGGCGTACCTACCGCAGCTCAAAGCTGGGTGCATTTTTGCCAAGTGCCTTACTGTTAT

CTTAGGACGGAAATCCACTATAAGATTATAGAAAGGAAGGCGGGCCGAGCGAATCGATTCAATTAAGTTATGTCACAAGG

GTGCTATAGCCTATTCCTAAGATTTGTACGTGCGTATGACTGGAATTAATAACCCCTCCCTGCACTGACCTTGACTGAAT

AACTGTGATACGACGCAAACTGAACGCTGCGGGTCCTTTATGACCACGGATCACGACCGCTTAAGACCTGAGTTGGAGTT

GATACATCCGGCAGGCAGCCAAATCTTTTGTAGTTGAGACGGATTGCTAAGTGTGTTAACTAAGACTGGTATTTCCACTA

GGACCACGCTTACATCAGGTCCCAAGTGGACAACGAGTCCGTAGTATTGTCCACGAGAGGTCTCCTGATTACATCTTGAA

GTTTGCGACGTGTTATGCGGATGAAACAGGCGGTTCTCATACGGTGGGGCTGGTAAACGAGTTCCGGTCGCGGAGATAAC

TGTTGTGATTGGCACTGAAGTGCGAGGTCTTAAACAGGCCGGGTGTACTAACCCAAAGACCGGCCCAGCGTCAGTGA...

Application: Discovering DNA motifs

...TTGGAACAACCATGCACGGTTGATTCGTGCCTGTGACCGCGCGCCTCACACGGAAGACGCAGCCACCGGTTGTGATG

TCATA**GGGAATTCCC**CATGTCGTGAATAATGCCTCGAATGATGAGTAATAGTAAAACGCAGGGGAGGTTCTTCAGTAGTA

TCAATATGAGACACATACAAACGGGCGTACCTACCGCAGCTCAAAGCTGGGTGCATTTTTGCCAAGTGCCTTACTGTTAT

CTTAGGACGGAAATCCACTATAAGATTATAGAAAGGAAGGCGGGCCGAGCGAATCGATTCAATTAAGTTATGTCACAAGG

GTGCTATAGCCTATTCCTAAGATTTGTACGTGCGTATGACTGGAATTAATAACCCCTCCCTGCACTGACCTTGACTGAAT

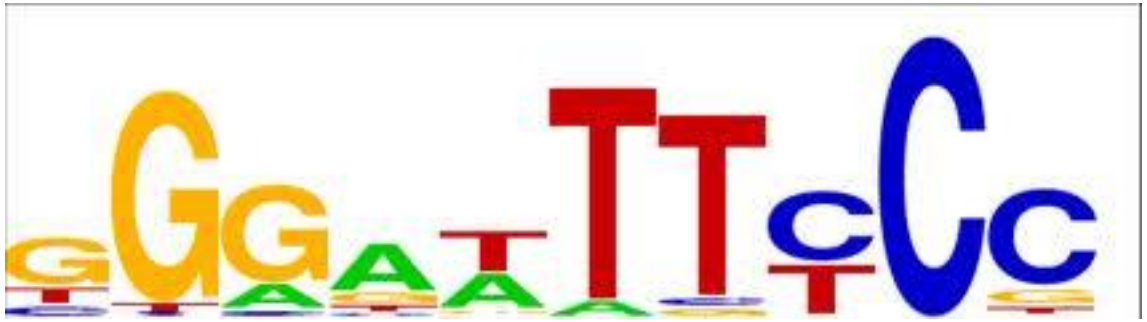
AACTGTGATACGACGCAAACTGAACGCTGCGGGTCCTTTATGACCACGGATCACGACCGCTTAAGACCTGAGTTGGAGTT

GATACATCCGGCAGGCAGCCAAATCTTTTGTAGTTGAGACGGATTGCTAAGTGTGTTAACTAAGAC**TGGTATTTCC**ACTA

GGACCACGCTTACATCAGGTCCCAAGTGGACAACGAGTCCGTAGTATTGTCCACGAGAGGTCTCCTGATTACATCTTGAA

GTTTGCGACGTGTTATGCGGATGAAACAGGCGGTTCTCATACGGTGGGGCTGGTAAACGAGTTCCGGTCGCGGAGATAAC

TGTTGTGATTGGCACTGAAGTGCGAGGTCTTAAACAGGCCGGGTGTACTAACCCAAAGACCGGCCCAGCGTCAGTGA...

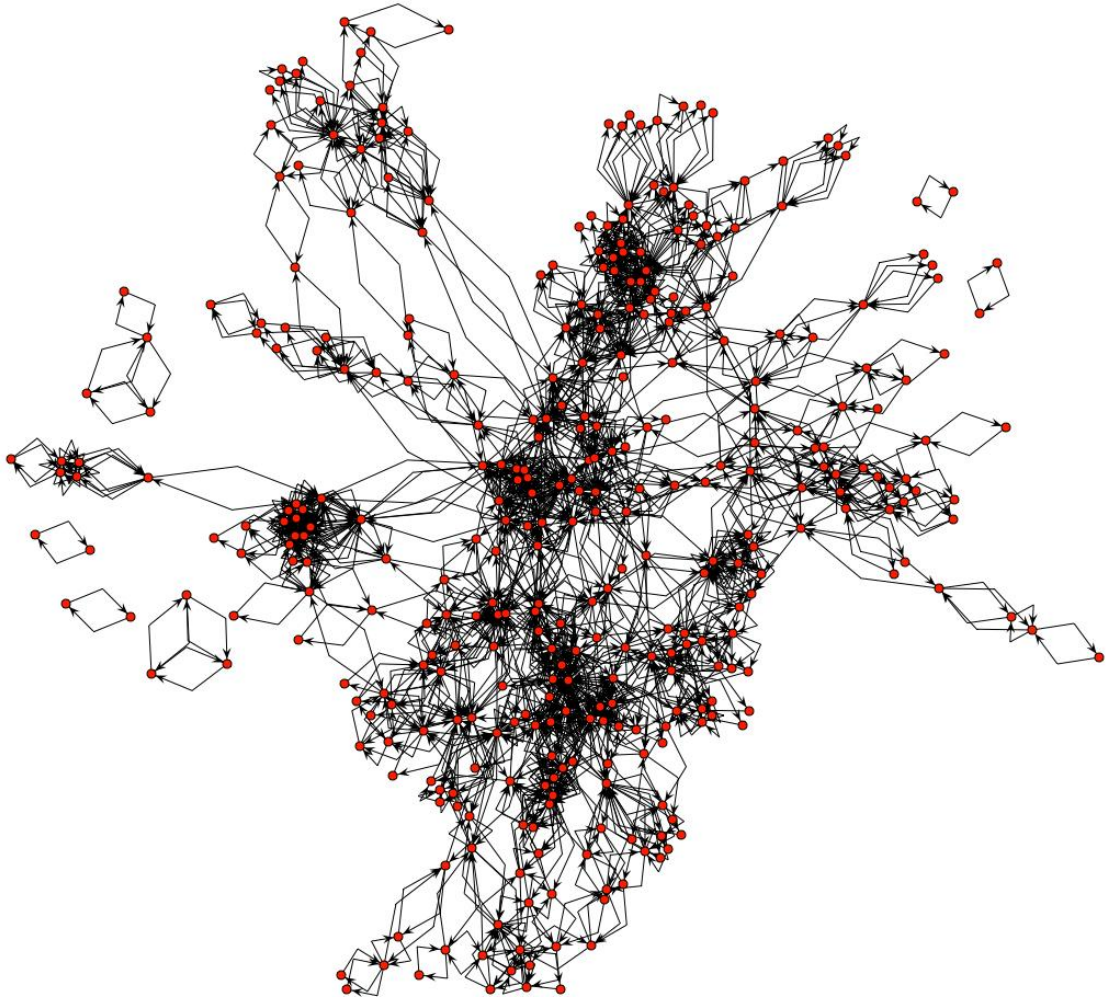


Application: User website behavior from clickstream data (from P. Smyth, UCI)

128.195.36.195, -, 3/22/00, 10:35:11, W3SVC, SRVR1, 128.200.39.181, 781, 363, 875, 200, 0, GET, /top.html, -,
128.195.36.195, -, 3/22/00, 10:35:16, W3SVC, SRVR1, 128.200.39.181, 5288, 524, 414, 200, 0, POST, /spt/main.html, -,
128.195.36.195, -, 3/22/00, 10:35:17, W3SVC, SRVR1, 128.200.39.181, 30, 280, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.195.36.101, -, 3/22/00, 16:18:50, W3SVC, SRVR1, 128.200.39.181, 60, 425, 72, 304, 0, GET, /top.html, -,
128.195.36.101, -, 3/22/00, 16:18:58, W3SVC, SRVR1, 128.200.39.181, 8322, 527, 414, 200, 0, POST, /spt/main.html, -,
128.195.36.101, -, 3/22/00, 16:18:59, W3SVC, SRVR1, 128.200.39.181, 0, 280, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.200.39.17, -, 3/22/00, 20:54:37, W3SVC, SRVR1, 128.200.39.181, 140, 199, 875, 200, 0, GET, /top.html, -,
128.200.39.17, -, 3/22/00, 20:54:55, W3SVC, SRVR1, 128.200.39.181, 17766, 365, 414, 200, 0, POST, /spt/main.html, -,
128.200.39.17, -, 3/22/00, 20:54:55, W3SVC, SRVR1, 128.200.39.181, 0, 258, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.200.39.17, -, 3/22/00, 20:55:07, W3SVC, SRVR1, 128.200.39.181, 0, 258, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.200.39.17, -, 3/22/00, 20:55:36, W3SVC, SRVR1, 128.200.39.181, 1061, 382, 414, 200, 0, POST, /spt/main.html, -,
128.200.39.17, -, 3/22/00, 20:55:36, W3SVC, SRVR1, 128.200.39.181, 0, 258, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.200.39.17, -, 3/22/00, 20:55:39, W3SVC, SRVR1, 128.200.39.181, 0, 258, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.200.39.17, -, 3/22/00, 20:56:03, W3SVC, SRVR1, 128.200.39.181, 1081, 382, 414, 200, 0, POST, /spt/main.html, -,
128.200.39.17, -, 3/22/00, 20:56:04, W3SVC, SRVR1, 128.200.39.181, 0, 258, 111, 404, 3, GET, /spt/images/bk1.jpg, -,
128.200.39.17, -, 3/22/00, 20:56:33, W3SVC, SRVR1, 128.200.39.181, 0, 262, 72, 304, 0, GET, /top.html, -,
128.200.39.17, -, 3/22/00, 20:56:52, W3SVC, SRVR1, 128.200.39.181, 19598, 382, 414, 200, 0, POST, /spt/main.html, -,

User 1	2 3 2 2 3 3 3 1 1 1 3 1 3 3 3 3
User 2	3 3 3 1 1 1
User 3	7 7 7 7 7 7 7 7
User 4	1 5 1 1 1 5 1 5 1 1 1 1 1 1
User 5	5 1 1 5
…	…

Application: social network analysis

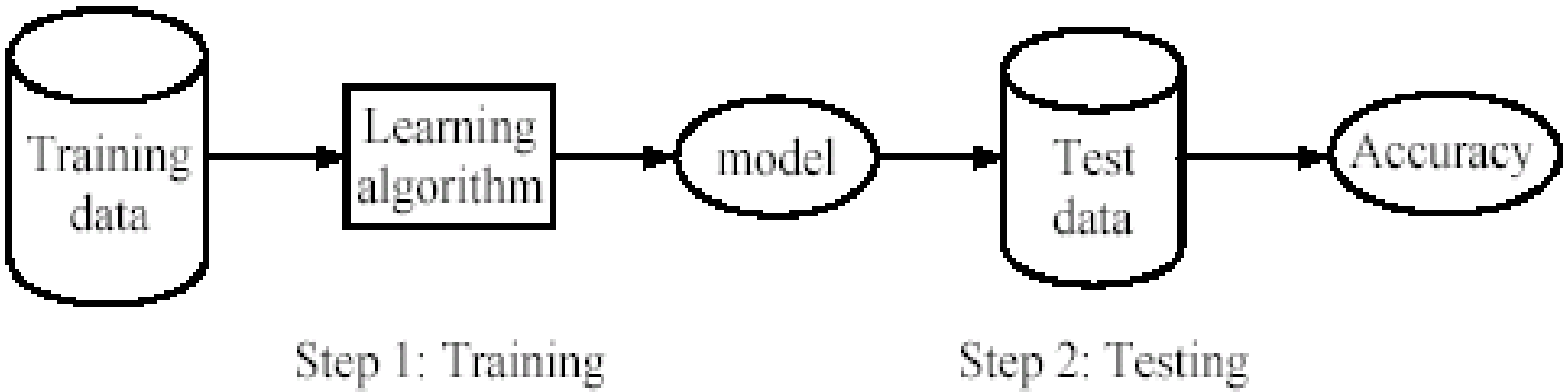


HP Labs email data
500 users, 20k connections
evolving over time

Supervised learning process: two steps

- Learning (training): Learn a model using the training data
- Testing: Test the model using unseen test data to assess the model accuracy

$$Accuracy = \frac{\text{Number of correct classifica tions}}{\text{Total number of test cases}},$$



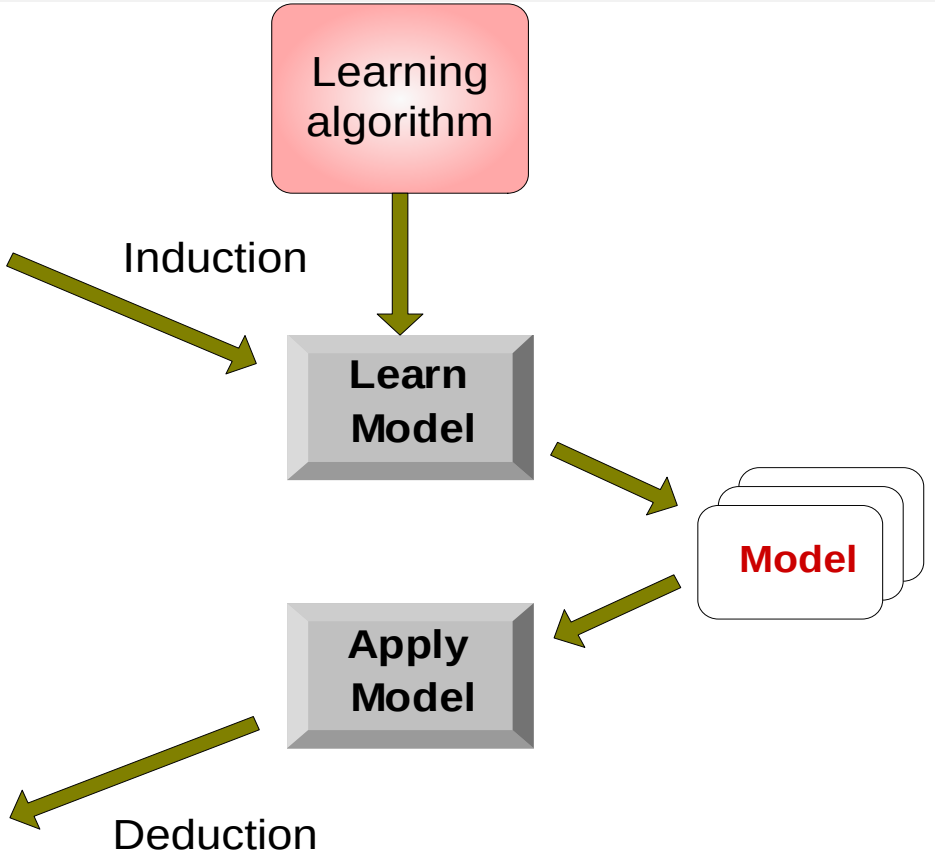
Illustrating Classification Task

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



What do we mean by learning?

- Given
  - a data set *D*,
  - a task *T,* and
  - a performance measure *M*,

  a computer system is said to **learn** from *D* to perform the task *T* if after learning the system's performance on *T* improves as measured by *M*.
- In other words, the learned model helps the system to perform *T* better as compared to no learning.

An example

- Data: Loan application data
- Task: Predict whether a loan should be approved or not.
- Performance measure: accuracy.

No learning: classify all future applications (test data) to the majority class (i.e., Yes):

Accuracy = 9/15 = 60%.

- We can do better than 60% with learning.

Fundamental assumption of learning

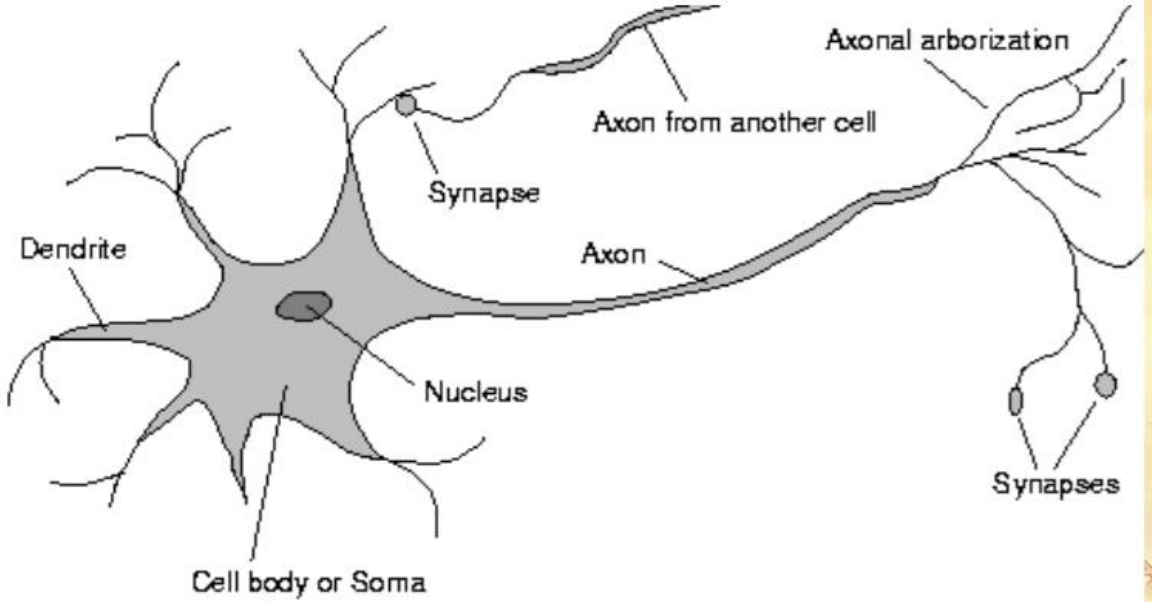
Assumption: The **distribution** of training examples is identical to the distribution of test examples (including future unseen examples).

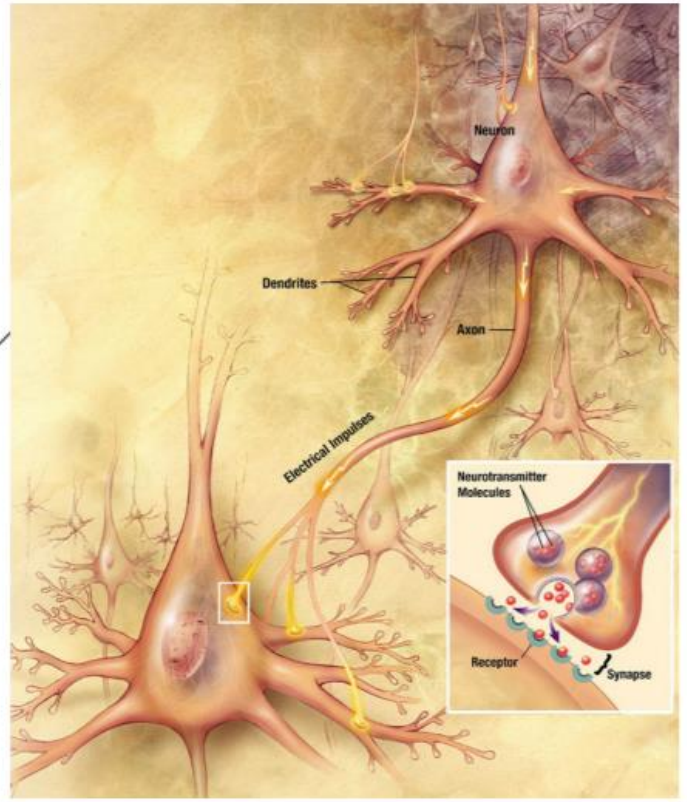
- In practice, this assumption is often violated to certain degree.
- Strong violations will clearly result in poor classification accuracy.
- To achieve good accuracy on the test data, training examples must be sufficiently representative of the test data.

Biological Inspirations

- Humans perform complex tasks like:
  - vision,
  - motor control,
  - or language understanding very well.
- One way to build intelligent machines is to try to imitate the (organizational principles of) human brain.

Biology of Neuron

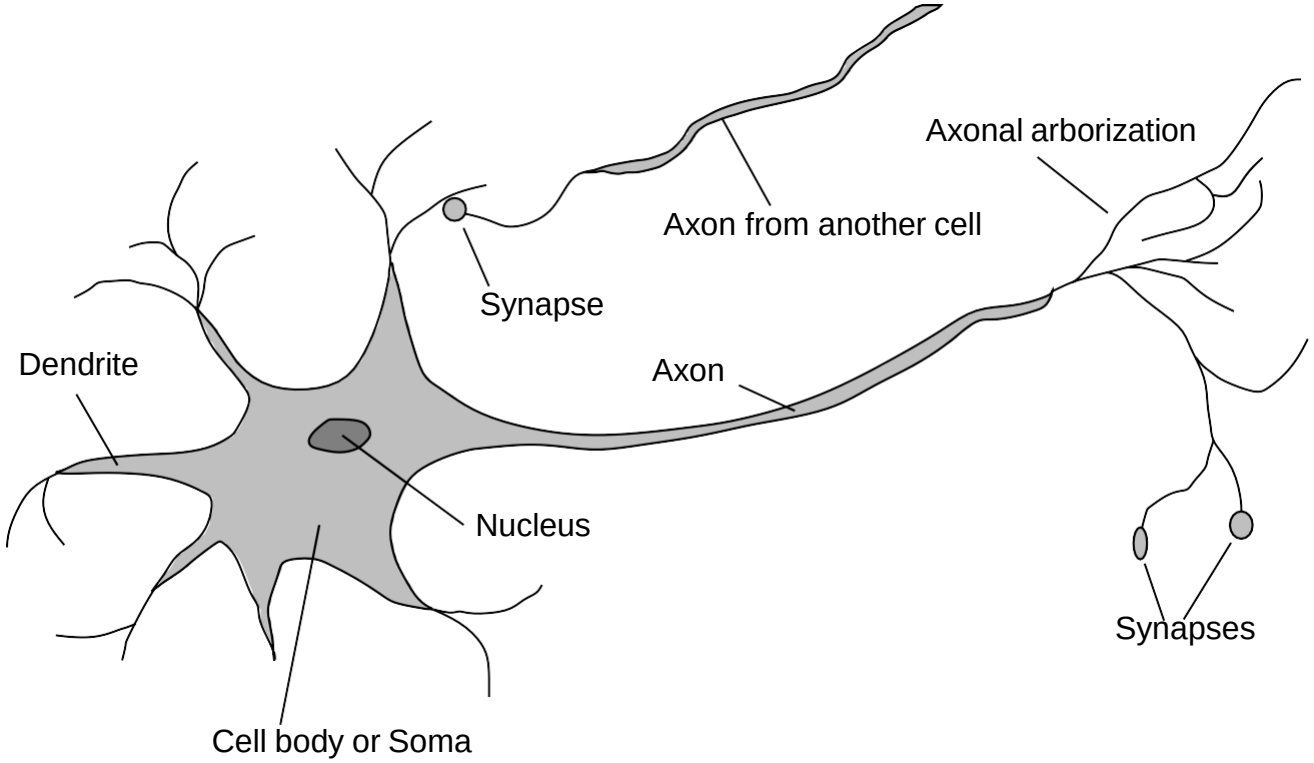




10^{11} neurons of > 20 types, 10^{14} synapses, 1ms–10ms cycle time
Signals are noisy “spike trains” of electrical potential

Biological Neuron

- dendrites: nerve fibres carrying electrical signals to the cell
- cell body: computes a non-linear function of its inputs
- axon: single long fiber that carries the electrical signal from the cell body to other neurons
- synapse: the point of contact between the axon of one cell and the dendrite of another, regulating a chemical connection whose strength affects the input to the cell.

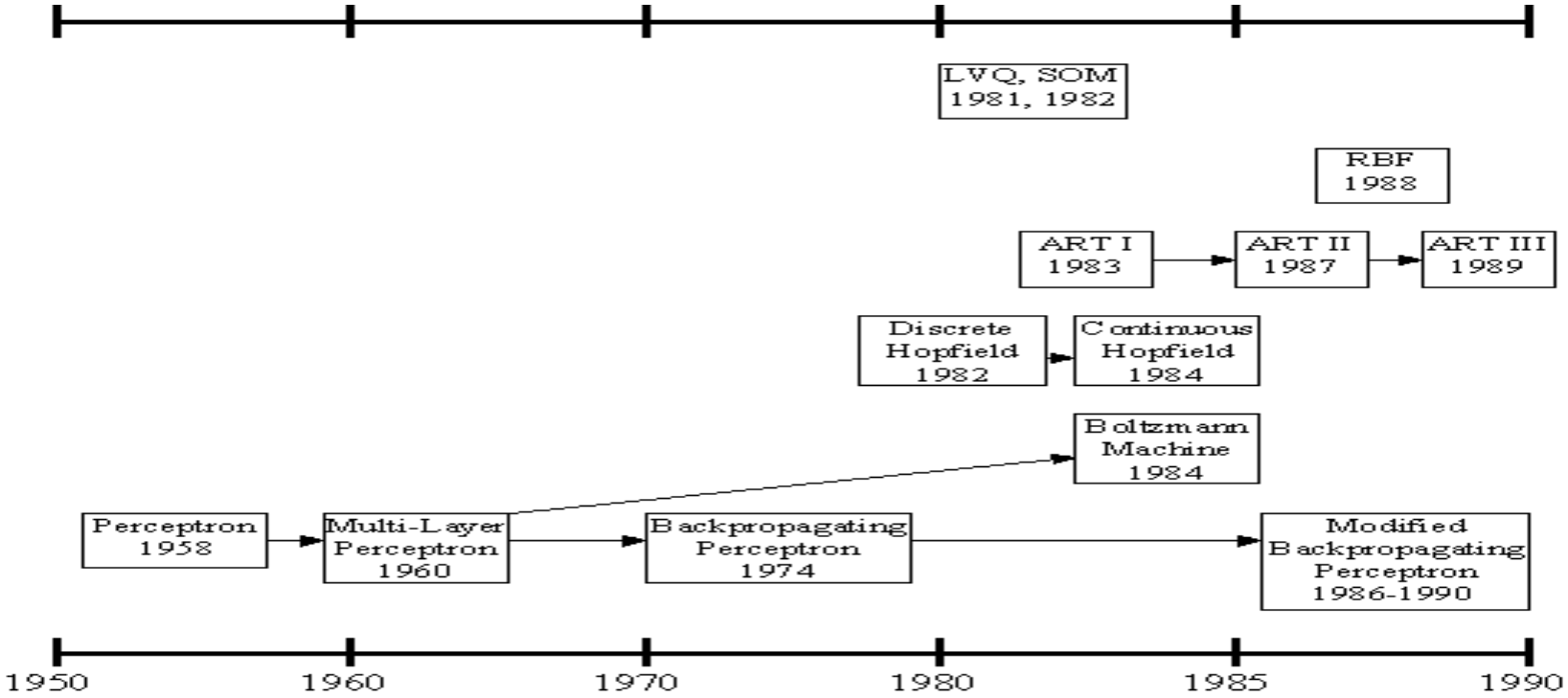


Artificial Neural Network

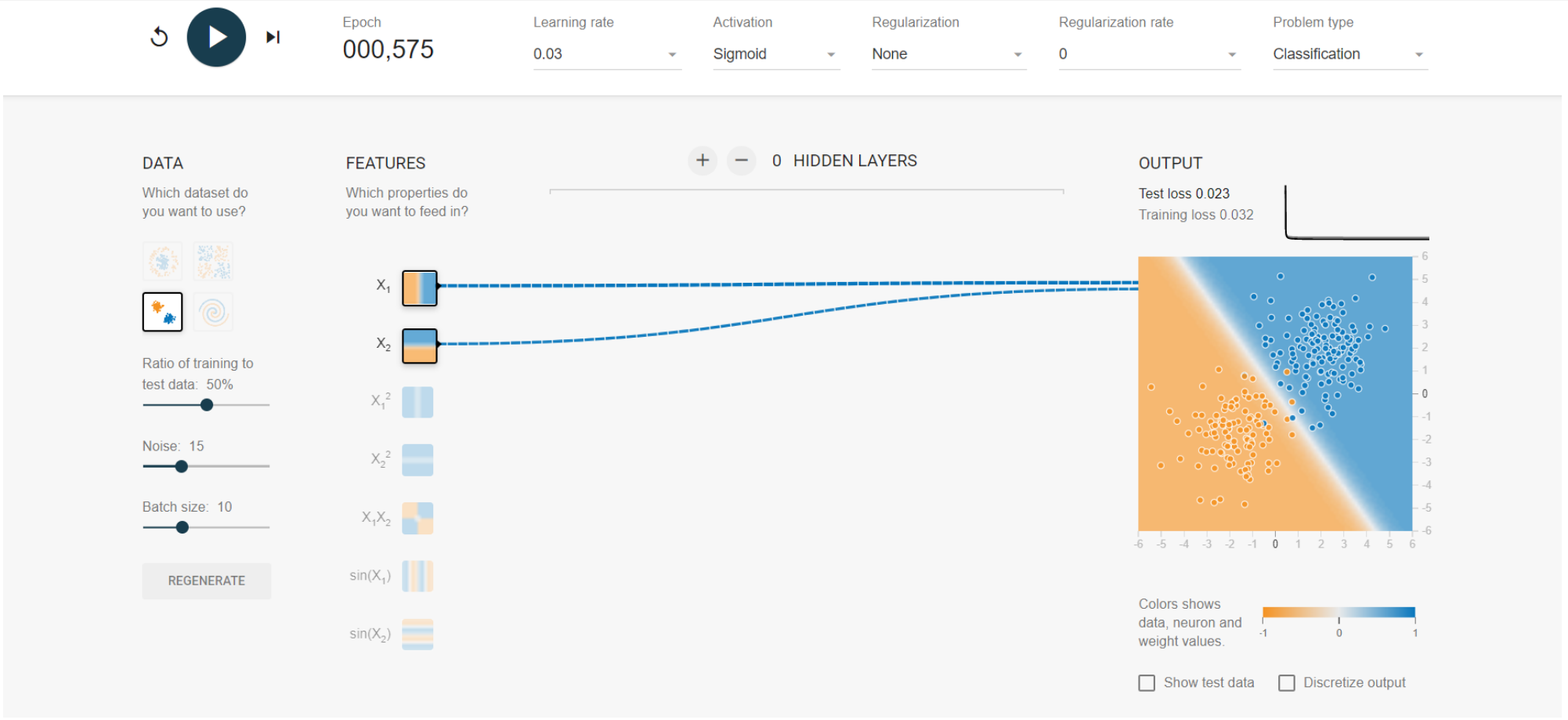
- Computational models inspired by the human brain:
  - Massively parallel, distributed system, made up of simple processing units (neurons)
  - Synaptic connection strengths among neurons are used to store the acquired knowledge.
  - Knowledge is acquired by the network from its environment through a learning process

History of ANN

- History of the ANNs stems from the 1940s, the decade of the first electronic computer.
- In 1957 Rosenblatt introduced the first concrete neural model, the perceptron.
- Rosenblatt also took part in constructing the first successful neurocomputer, the Mark I Perceptron.
- After this, the development of ANNs has proceeded as described in *Figure*.



Try It!!!



https://playground.tensorflow.org/