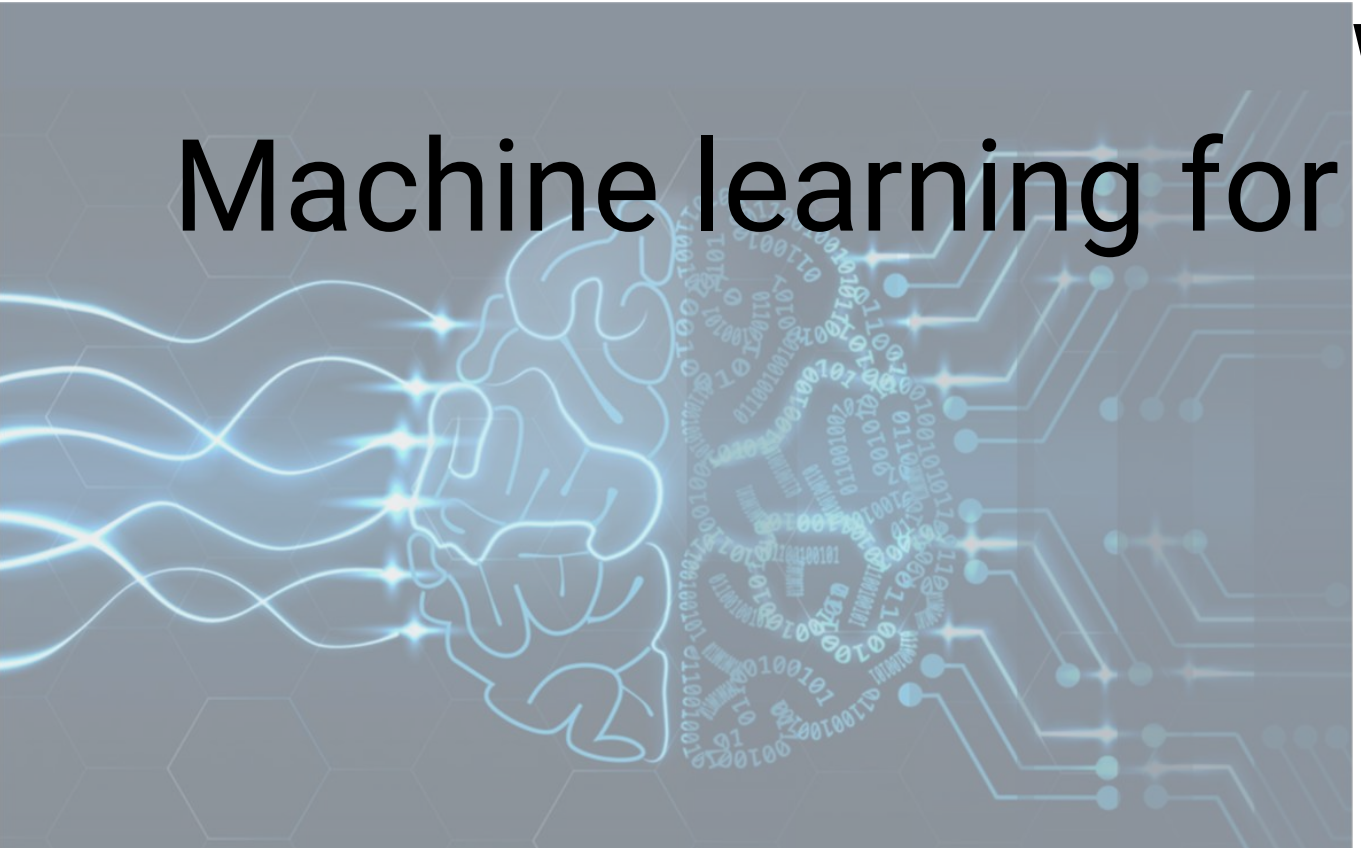


Welcome to: Machine learning for proteins

Parisa Hosseinzadeh

Spring 2022

Introduction

Name
Pronouns
Lab/year
What are you hoping to learn from this class
How do you evaluate your coding/biochemistry skills

If you want to recommend a book/movie, it would be … ?

Class core values

1. Be **respect**ful to yourself and others
2. Be **confident** and believe in yourself
3. Always do your **best**
4. Be **cooperative**
5. Be **creative**
6. Have **fun**
7. Be **patient** with yourself while you learn
8. Don’t be shy to **ask “stupid” questions**

About our class

What to bring?

- Yourself
- Your attention and curiosity
- Your laptops

About our class

What to bring?

- Yourself
- Your attention and curiosity
- Your laptops

What do I need to know?

- Basic biochemistry
- Basic learning
- Some python

About our class

Pre-class assessments (So you think you know X)

- Quiz to check your knowledge on the topic
- Short videos/readings that will bring you up to speed

Assessments are always due before the class

About our class

Pre-class assessments (So you think you know X)

- Quiz to check your knowledge on the topic
- Short videos/readings that will bring you up to speed

Post-class assignments

- Short quizzes
- Assay questions
- Readings (for journal club)
- Running a guided code
- …

About our class

Pre-class assessments (So you think you know X)

- Quiz to check your knowledge on the topic
- Short videos/readings that will bring you up to speed

Post-class assignments

- Short quizzes
- Assay questions
- Readings (for journal club)
- Running a guided code
- …

Assignments are always due the next Wednesdays, 3 PM

Activities

- Pre and post class assignments
- In class activities
  - Asking questions, submitting in-class work, presence
  - Either through canvas or google colab (gmail account)
  - 1 session allowed absence

Activities

Pre and post class assignments
In class activities
Journal presentation

slido.com #994666

Week 2	Karly
Week 3	Judah
Week 4	Lejla
Week 5	Noora
Week 6	Liza
Week 7	Alonso
Week 8	Cora
Week 9	Andrew

Activities

Pre and post class assignments
In class activities
Journal presentation
Final project
- Literature review
- Or
- Coding project

Potential coding projects

1. Predicting stability of designed proteins tested on the surface of yeast
2. Predicting whether a peptide is antimicrobial or not
3. Generating antimicrobial peptides
4. Generating new sequences for proteins of a known family

Potential review projects

1. Applications of deep learning in predicting protein stability
2. Applications of deep learning in protein design
3. Applications of machine learning in protein function prediction
4. Graph learning models of proteins

Let me know by next Monday

Activities

Pre and post class assignments
In class activities
Journal presentation
Final project
“Poster” session day
    Presenting your findings

Panels/Guests

Experts on certain areas of learning

- Dr. Mehanian for CNN

Panels/Guests

Experts on certain areas of learning
- Dr. Mehanian for CNN

Guest lectures on specialized topics
- Ethics of ML
- Application of ML

Feedback

Assignments and activities

- Detailed feedback provided by instructor
- Answering questions during hands-on sessions
- Help navigate the paper
- Office hours: 30 min after class each class

Feedback

Assignments and activities

- Detailed feedback provided by instructor
- Answering questions during hands-on sessions
- Help navigate the paper
- Office hours: 30 min after class each class

Feedback from you

- Mid-term and final feedback
- Personal feedback (email, anonymous feedback)
- 2 + 1 feedback every two weeks

About canvas

Spring 2022

- Home
- Announcements
- Assignments
- Grades
- Files
- Syllabus
- Chat
- Zoom Meetings
- UO Course Surveys

Machine learning for protein prediction and design

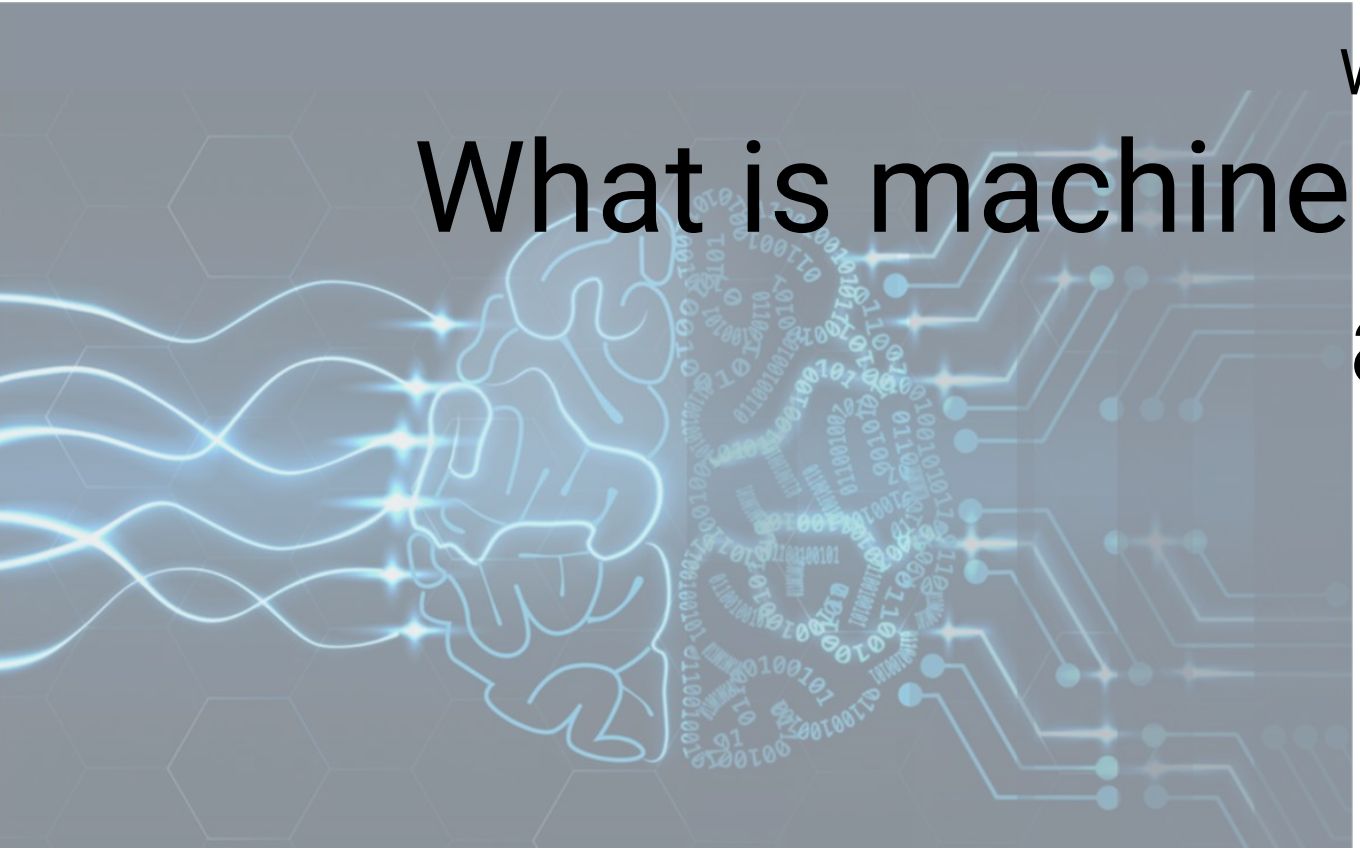
General information

Class times	Mondays/Wednesdays 3:00 - 4:50
Office hours	Mondays/Wednesdays half and hour after class
Instructor	Parisa Hosseinzadeh (*she/her*) -- email: **parisah 'at' uoregon.edu**
Location	KC 156, or if needed zoom (https://uoregon.zoom.us/j/99426388835)

Questions?

Week 1, Lecture 1

What is machine learning anyways

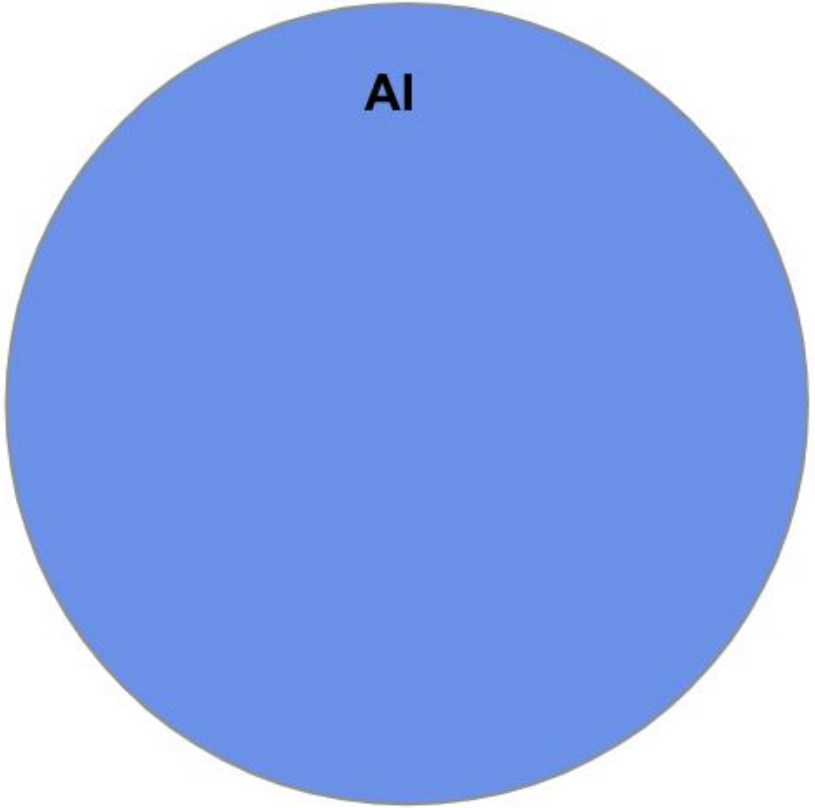


Learning Objectives

1. Describe the basic concept of machine learning
2. Explain the difference between supervised and unsupervised learning
3. Identify the difference between predictive and generative tasks
4. Calculate different performance metrics for a binary classification task
5. Clearly explain the concept of AUC-ROC

A gentle introduction to machine learning

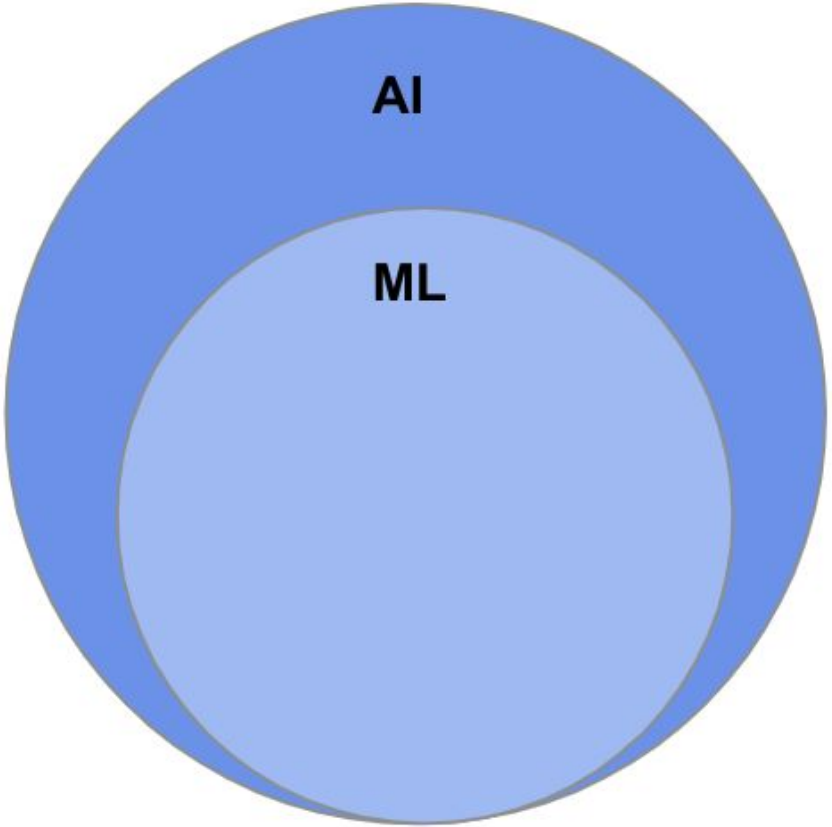
Artificial Intelligence



A gentle introduction to machine learning

Artificial Intelligence

Machine Learning

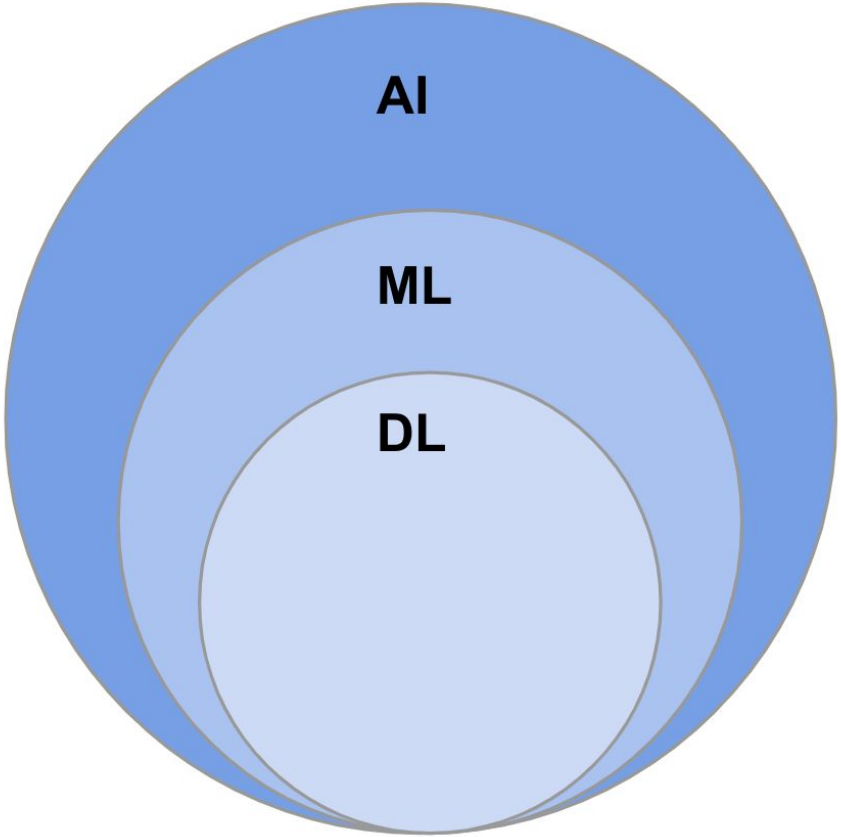


A gentle introduction to machine learning

Artificial Intelligence

Machine Learning

Deep learning



Putting it into [historic] perspective

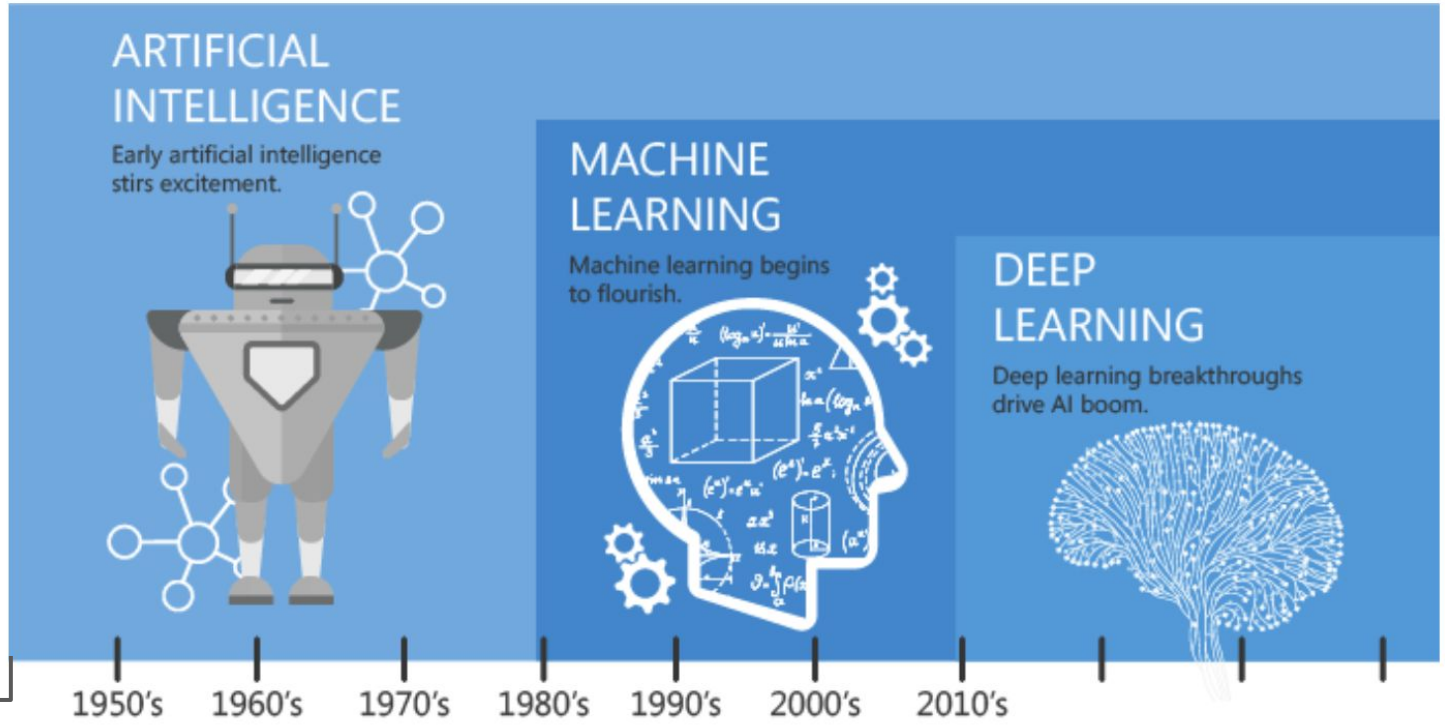
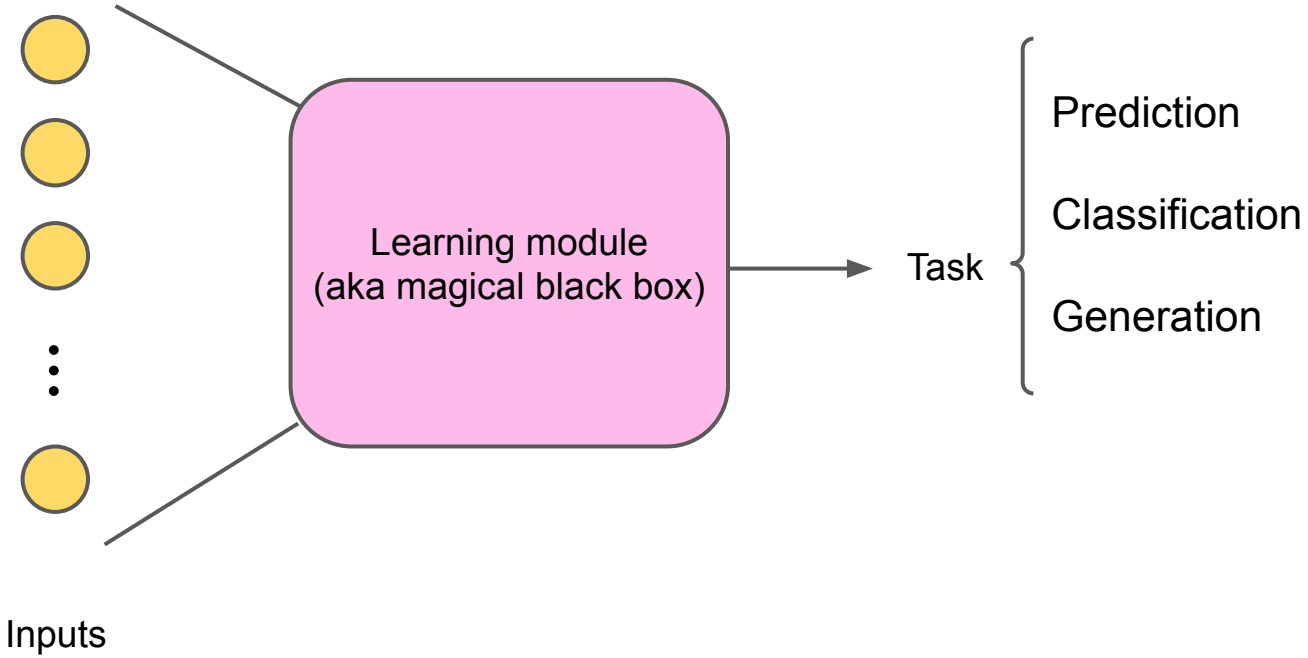
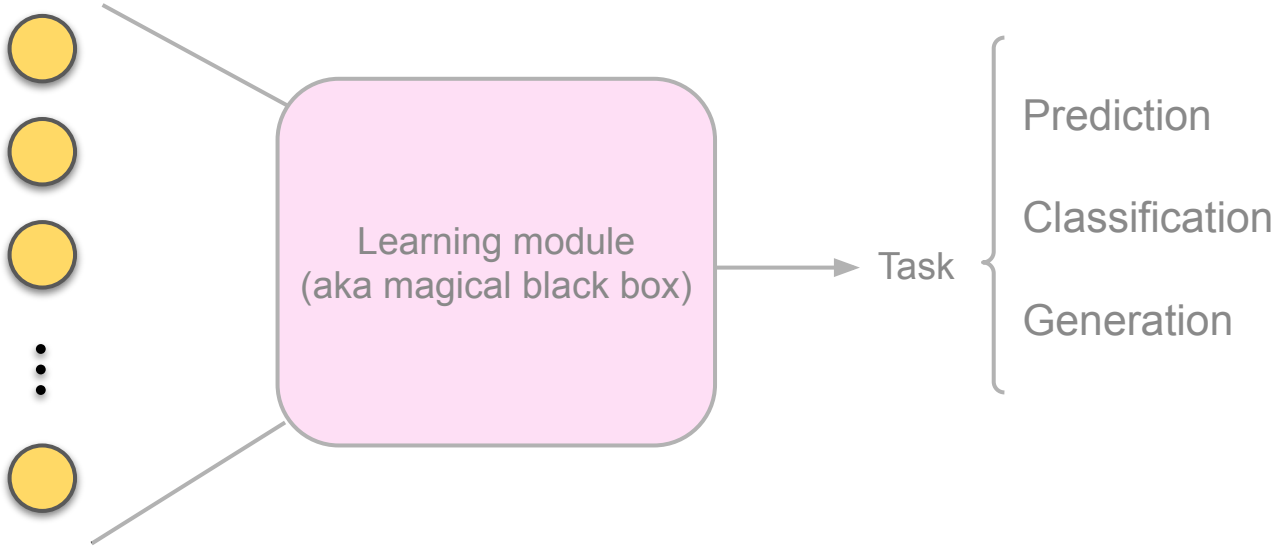


Image from tds – An introduction to ML for beginners

The basic components of a learning system

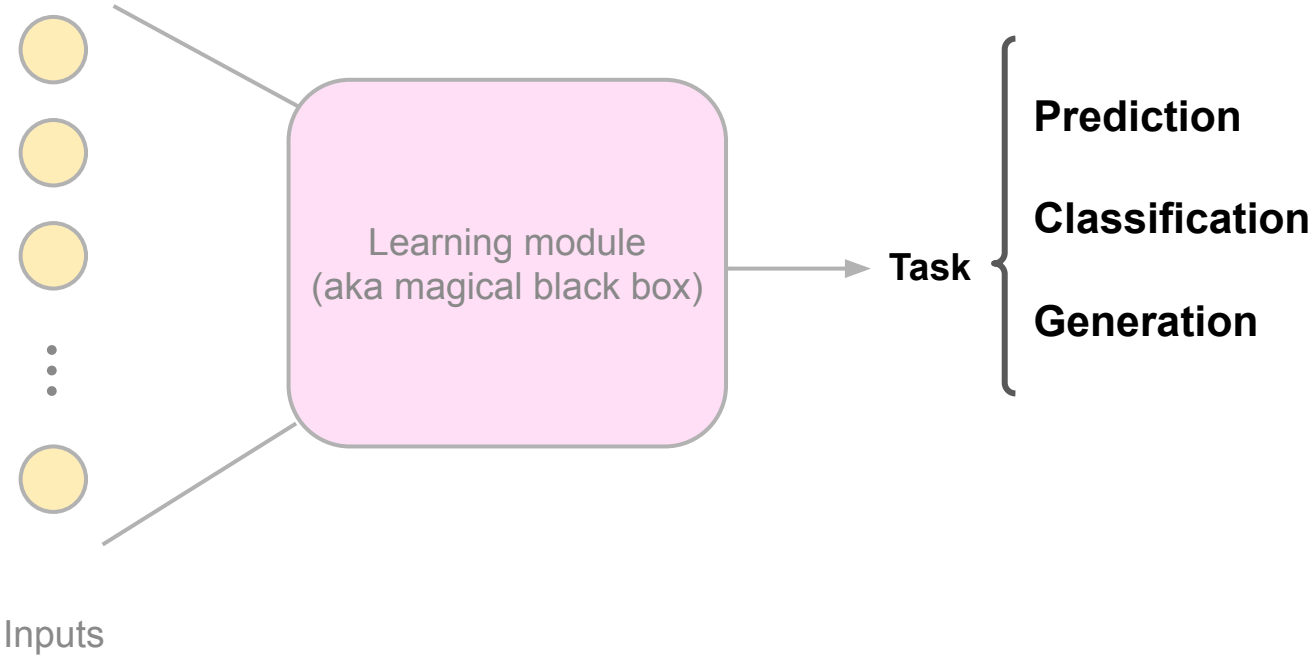


The basic components of a learning system

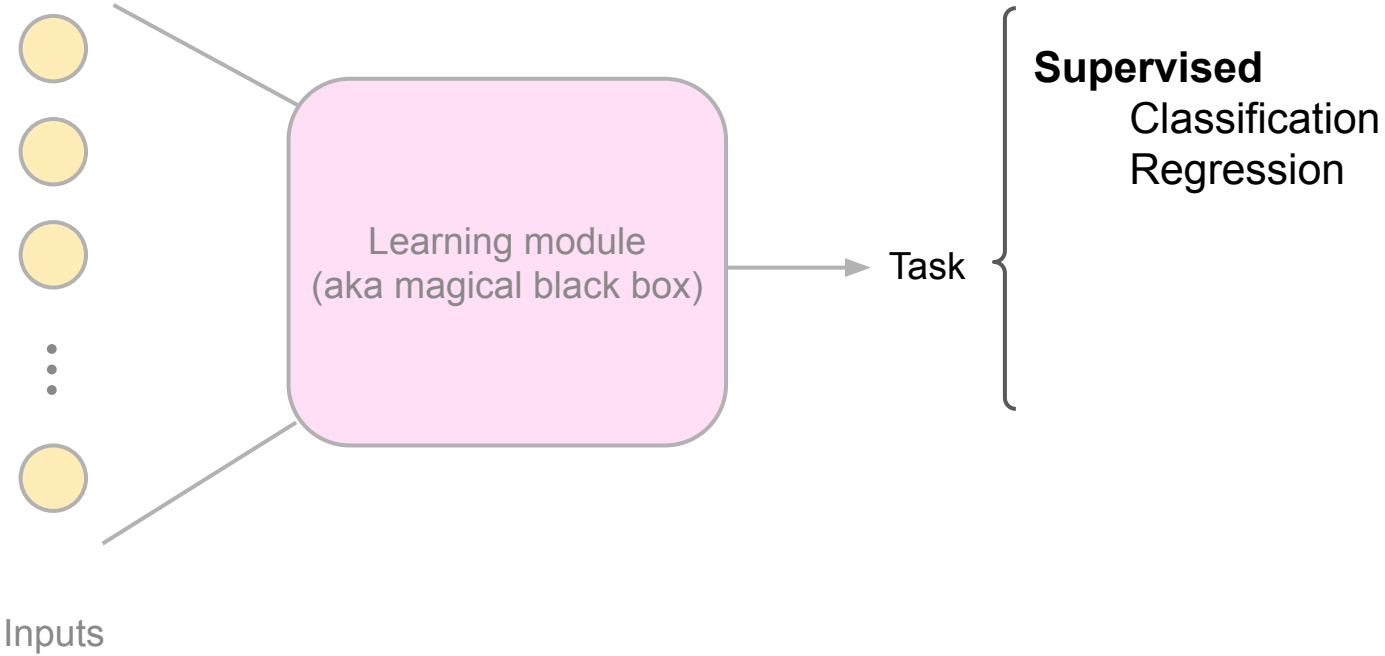


Inputs

The basic components of a learning system



Supervised tasks use labeled data



Supervised tasks use labeled data

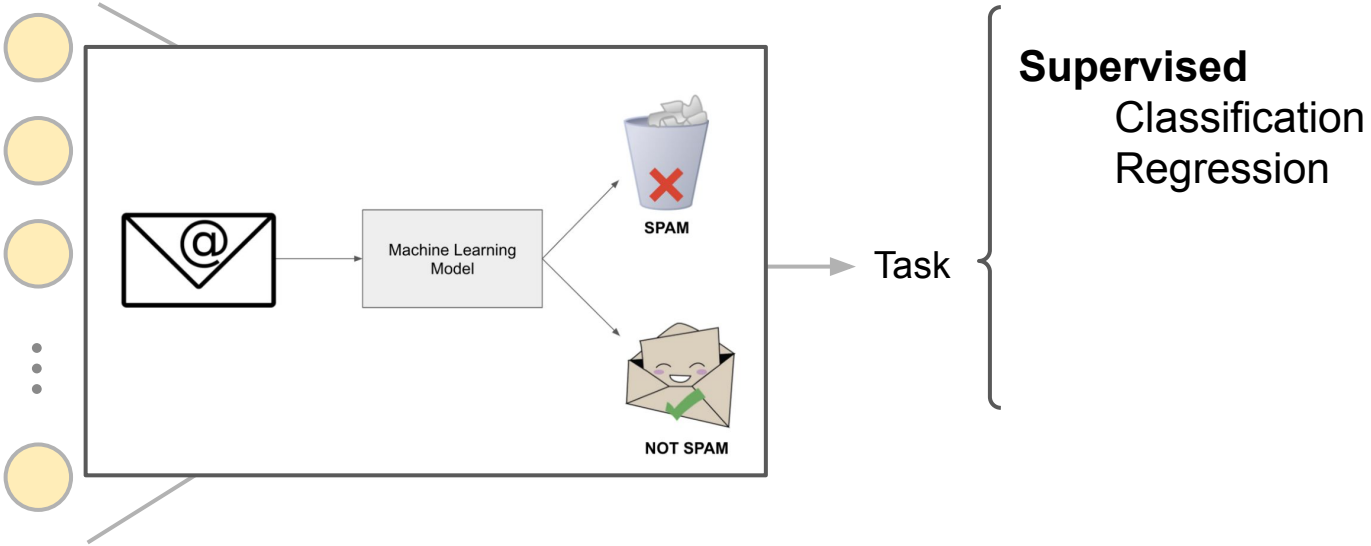
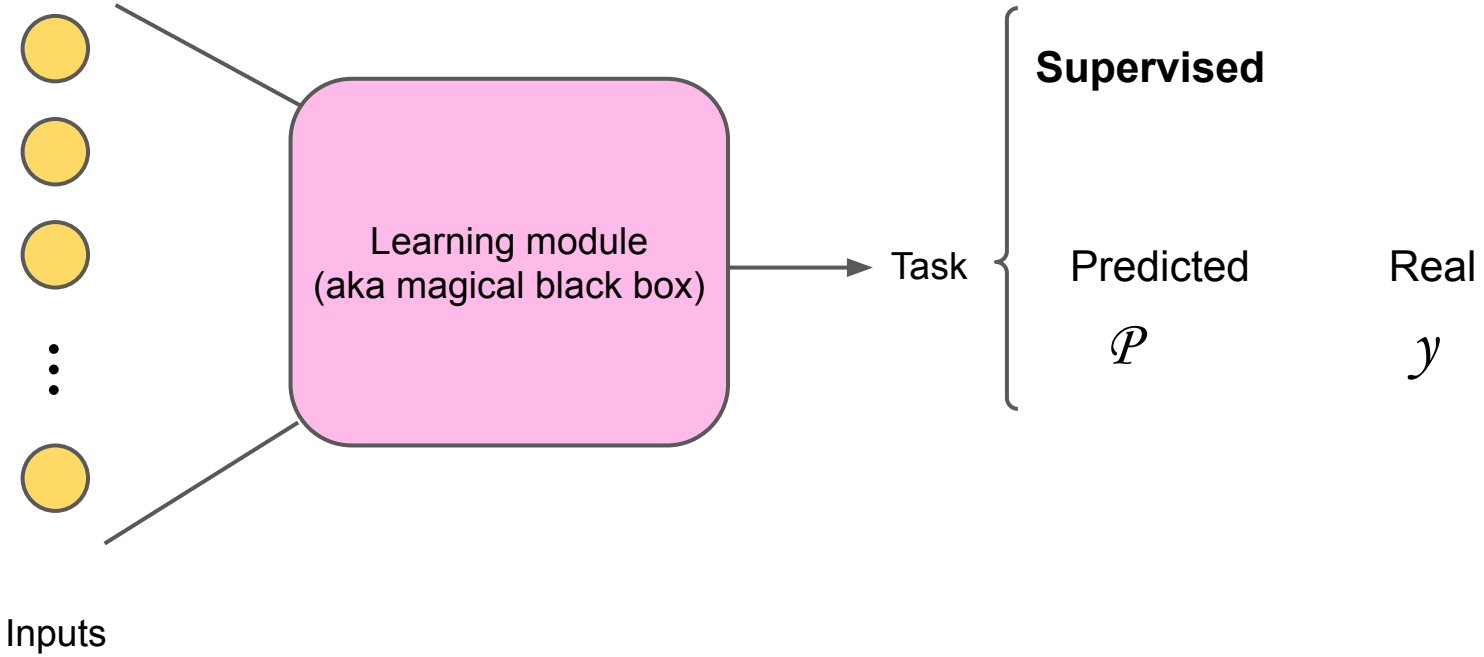
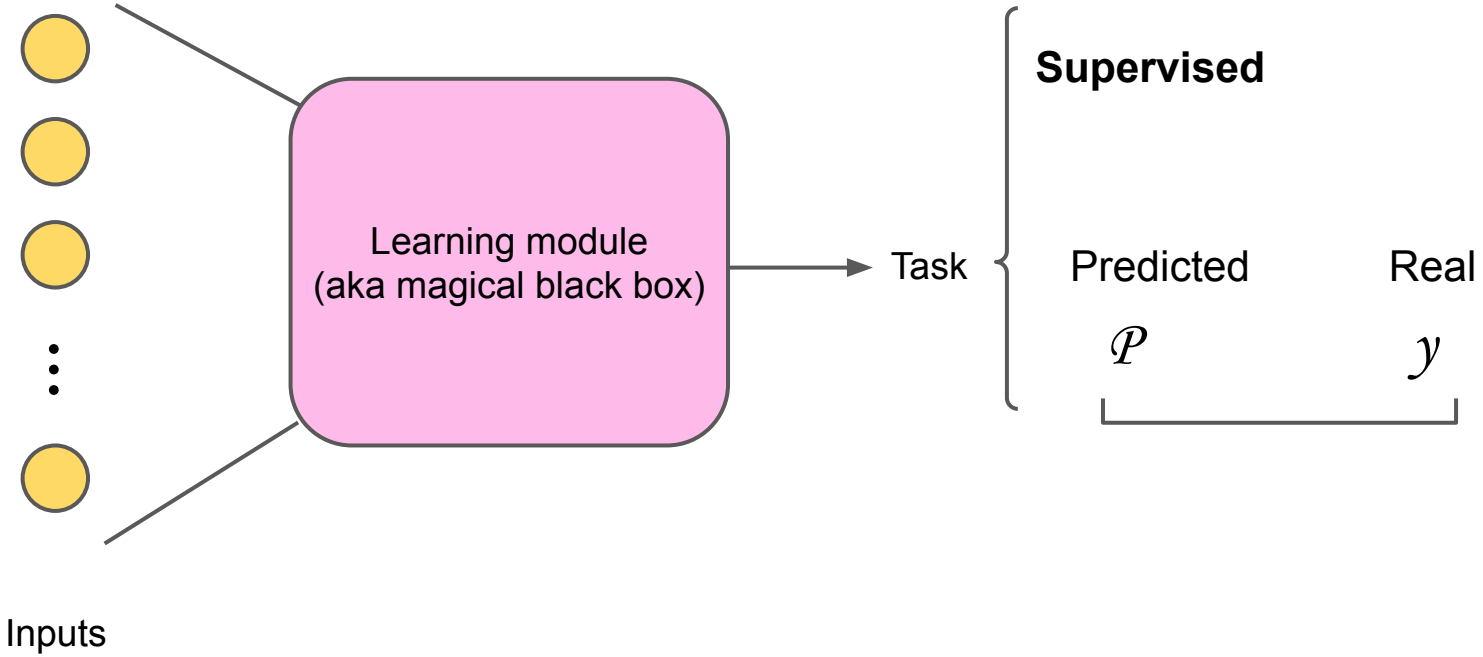


Image from tds: logic and implementation of spam filter

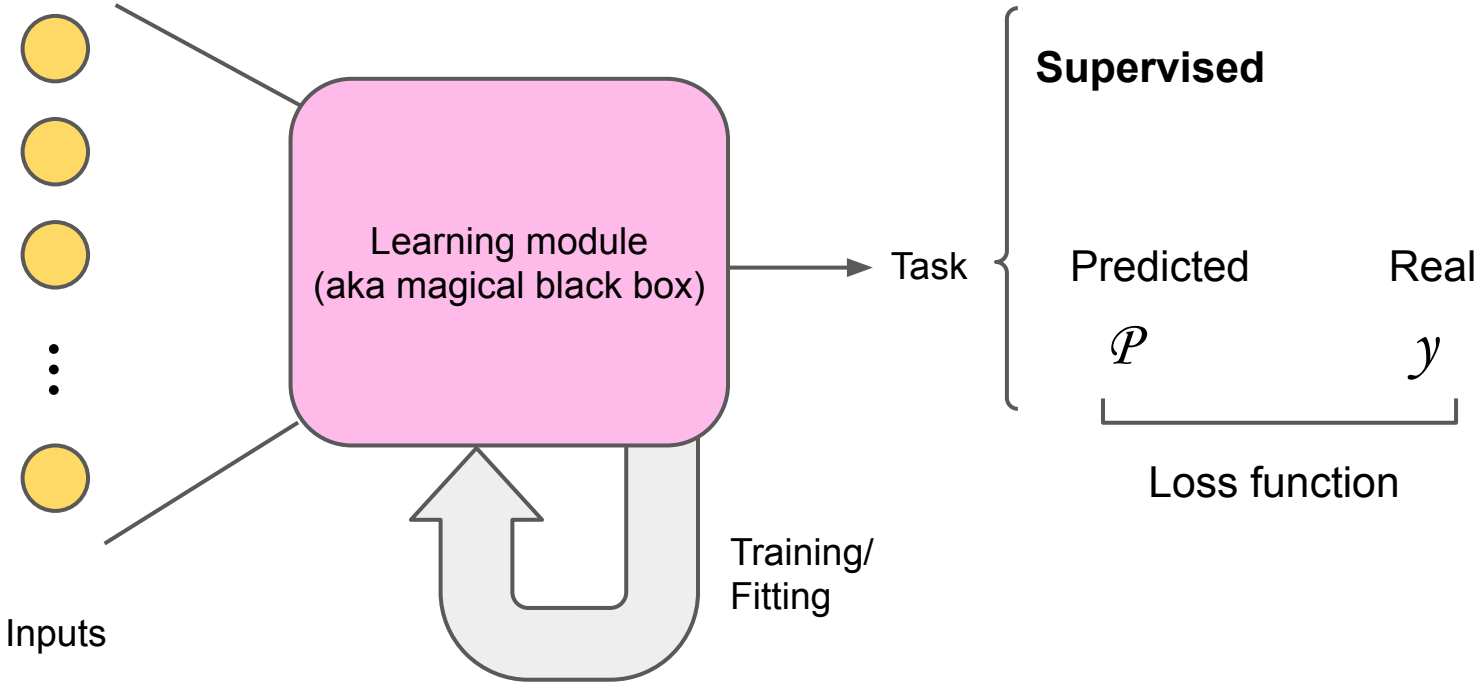
Supervised tasks use labeled data



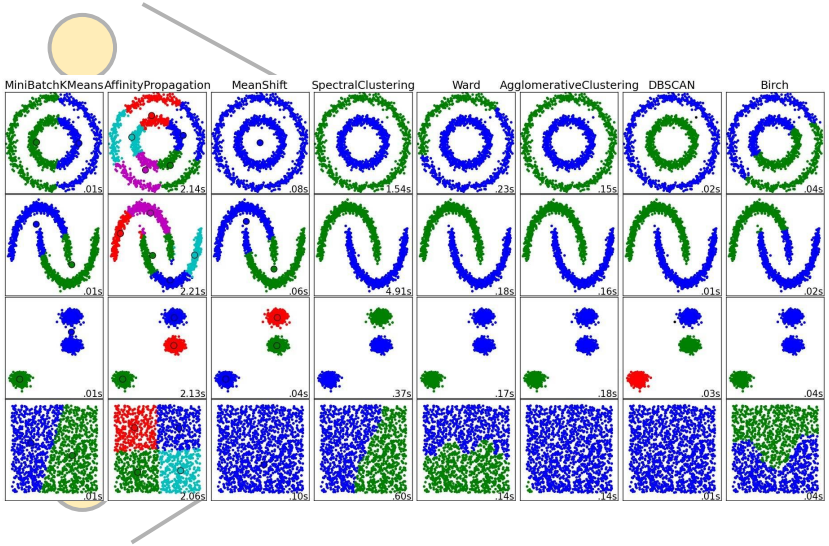
Supervised tasks use labeled data



Supervised tasks use labeled data



Unsupervised tasks do not need labeled data



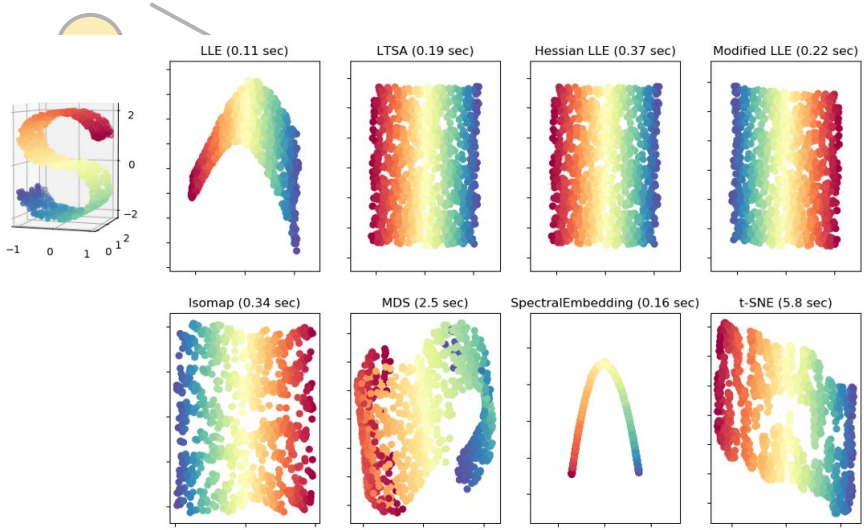
Inputs

→ Task

- Supervised
- **Unsupervised**
  - Clustering

Image from scikit learn – clustering

Unsupervised tasks do not need labeled data



Inputs

Task

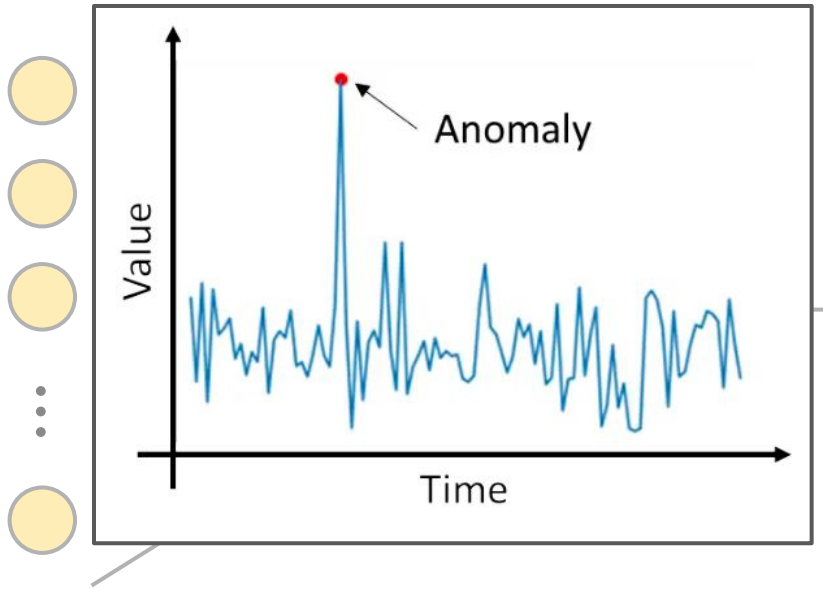
Supervised

Unsupervised

- Clustering
- Dimensionality reduction

Image from tds: Guide to dimensionality reduction in machine learning

Unsupervised tasks do not need labeled data



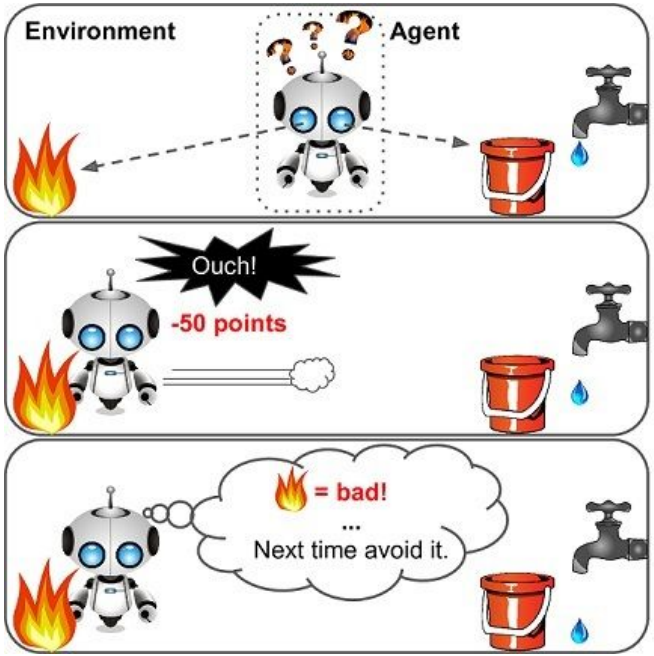
Inputs

Task

- Supervised
- **Unsupervised**
  - Clustering
  - Dimensionality reduction
  - Parametric unsupervised
  - Anomaly detection

Image from analytics vidhya: univariate anomaly detection

Unsupervised tasks do not need labeled data



1. Observe
2. Select action using policy
3. Action!
4. Get reward or penalty
5. Update policy (learning step)
6. Iterate until an optimal policy is found

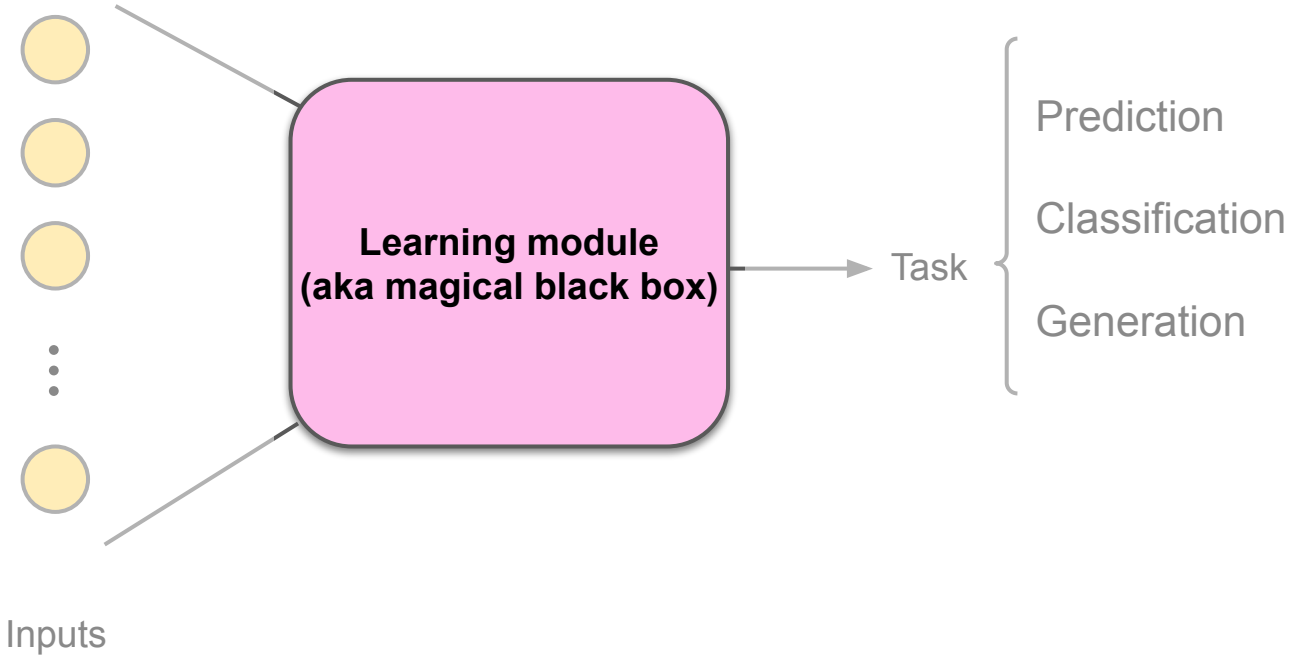
Inputs

Task

- Supervised
- Unsupervised
- **Reinforcement learning**
  - Learning task
  - Navigation
  - Skill acquisition
  - Real time decision

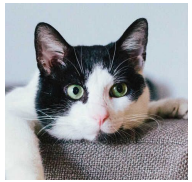
Image from tds – An introduction to ML for beginners

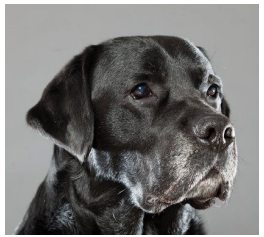
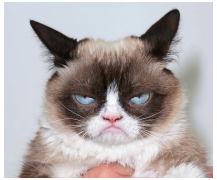
The basic components of a learning system



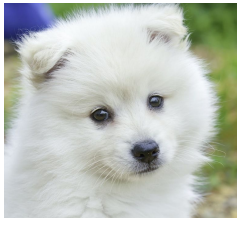
Measuring performance

Performance can be assessed by analyzing true positives and negatives

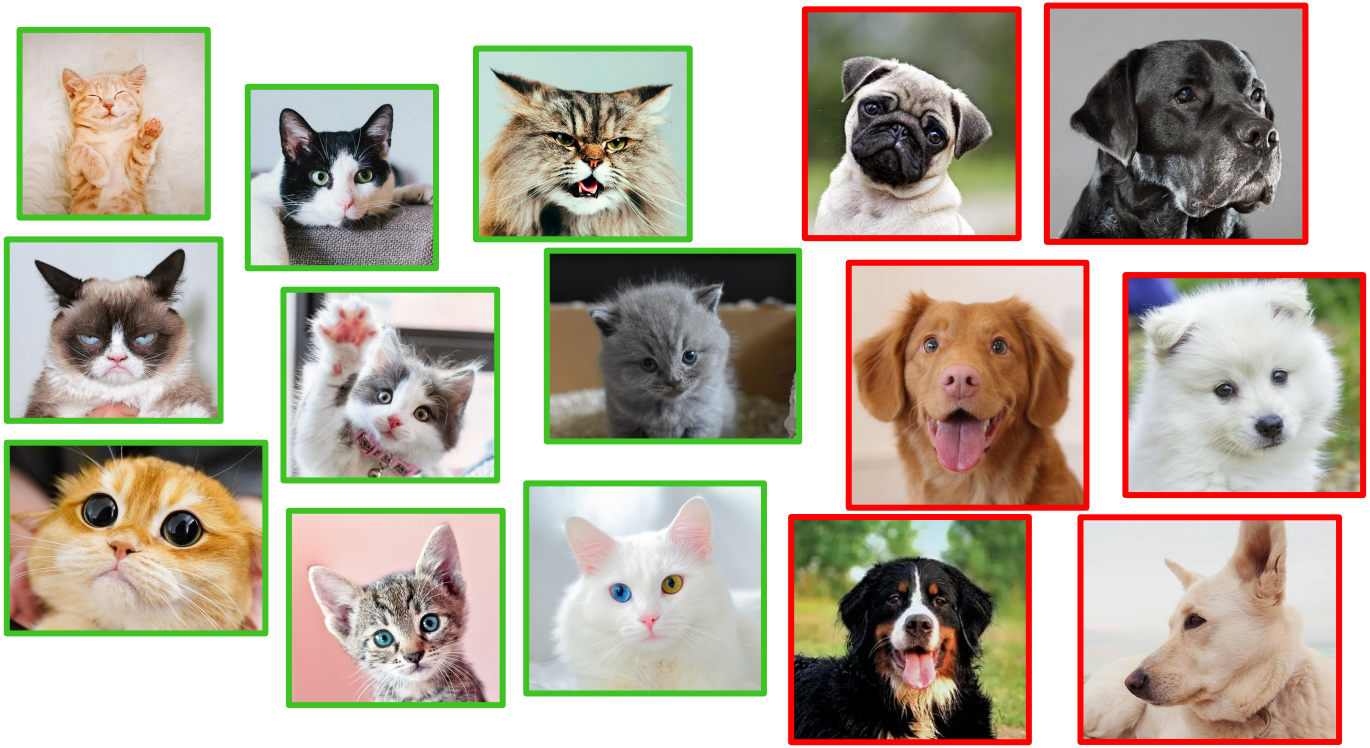






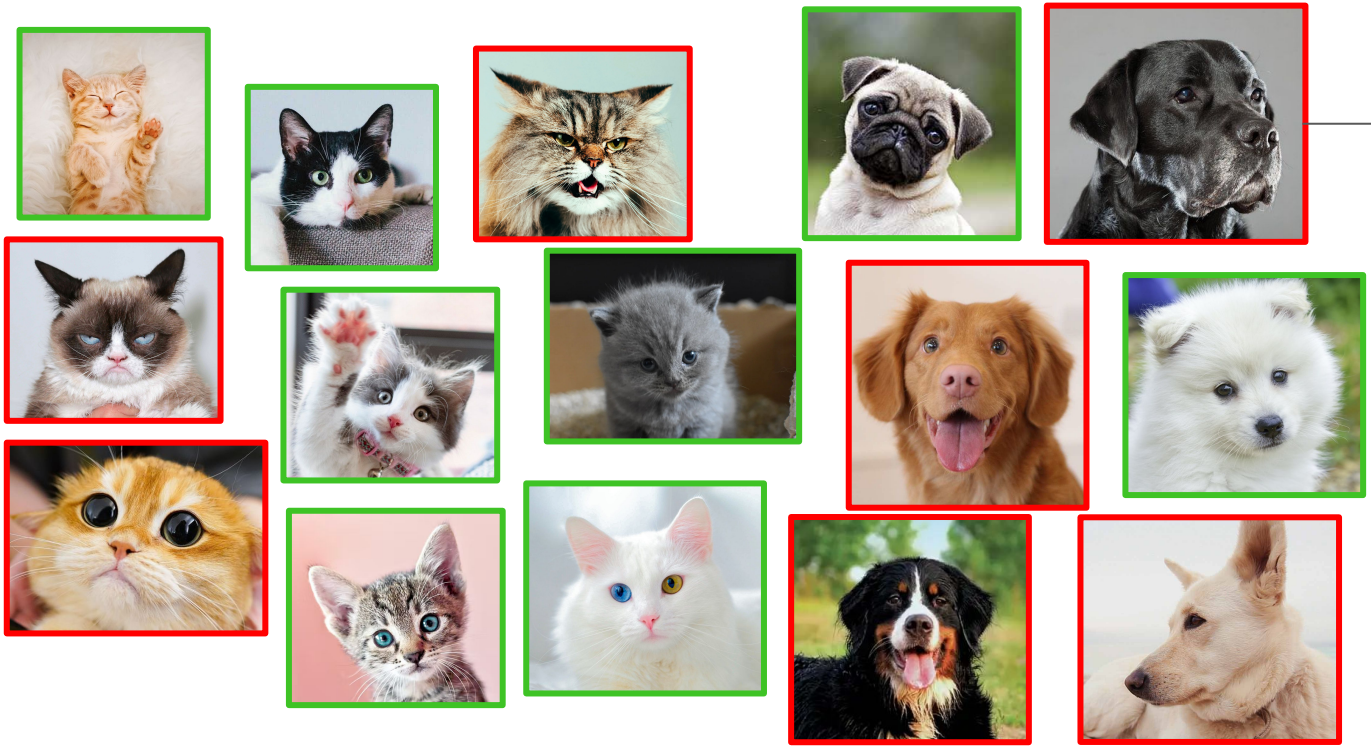

Images from various sources on the web

Performance can be assessed by analyzing true positives and negatives



Images from various sources on the web

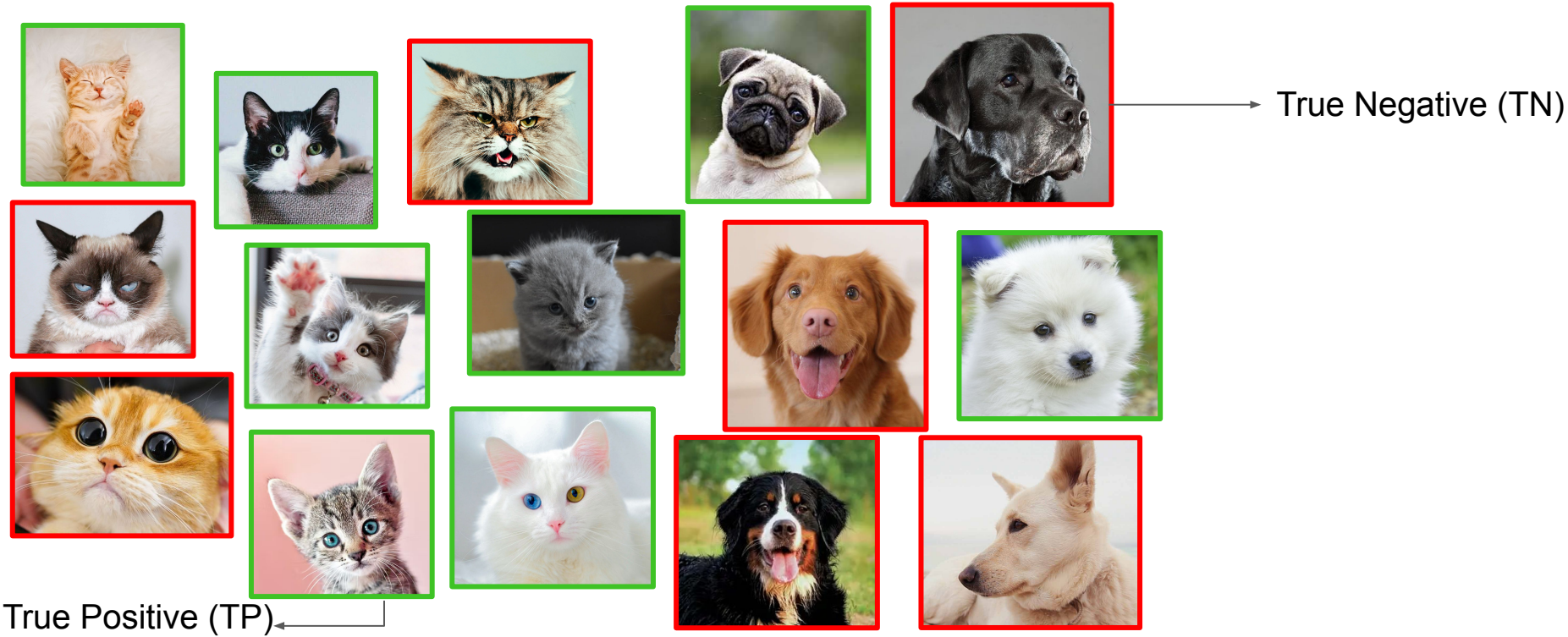
Performance can be assessed by analyzing true positives and negatives



True Negative (TN)

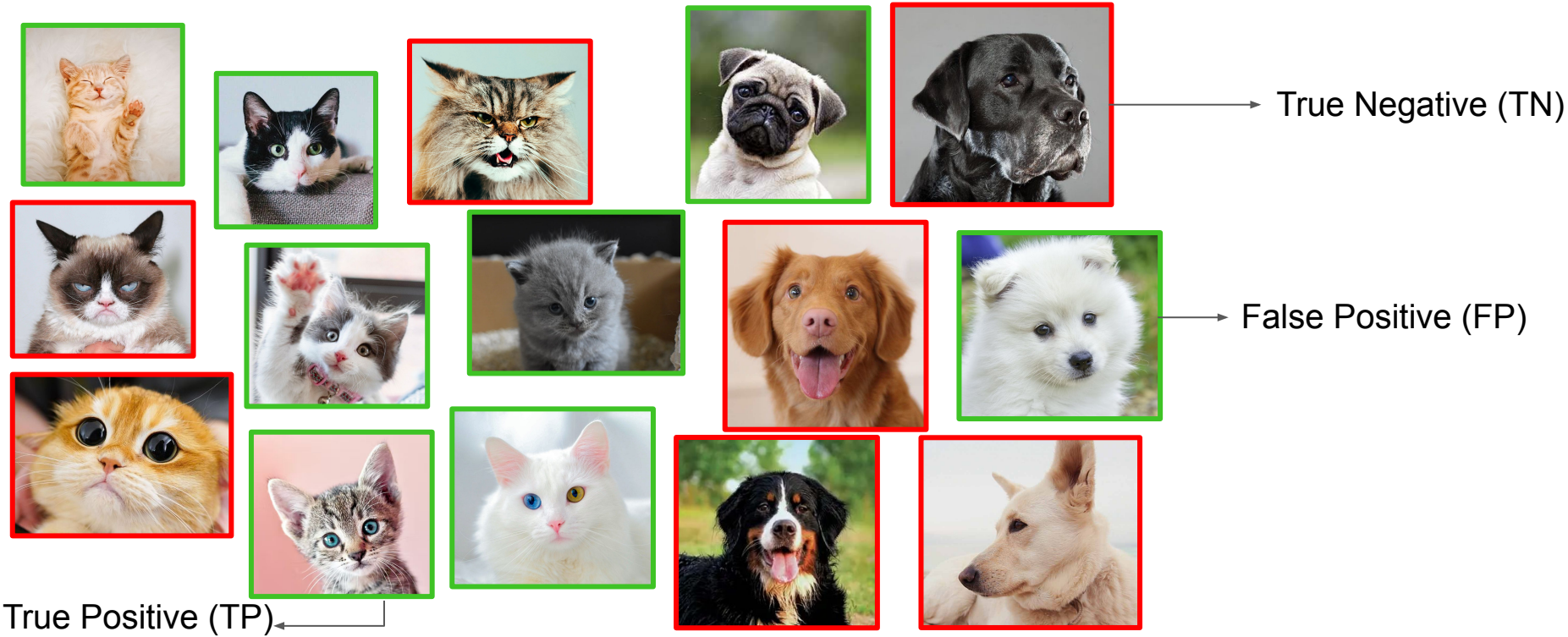
Images from various sources on the web

Performance can be assessed by analyzing true positives and negatives



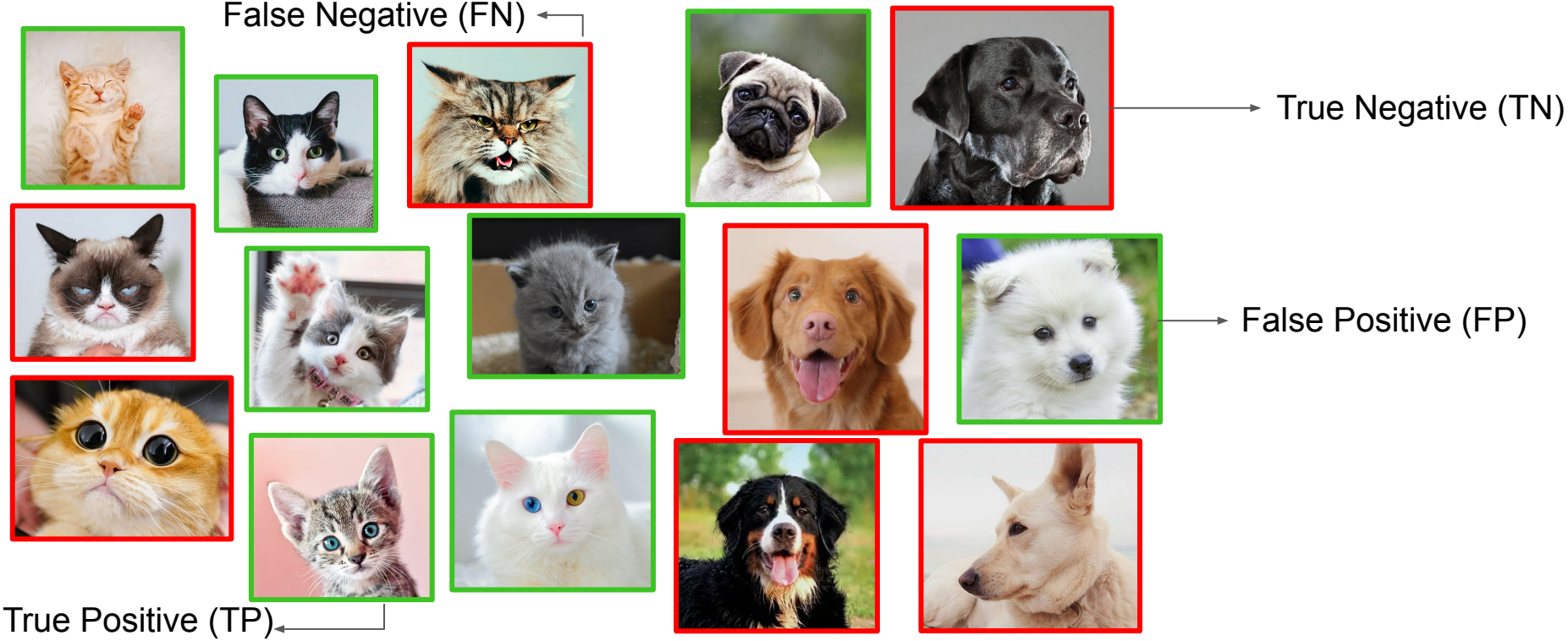
Images from various sources on the web

Performance can be assessed by analyzing true positives and negatives



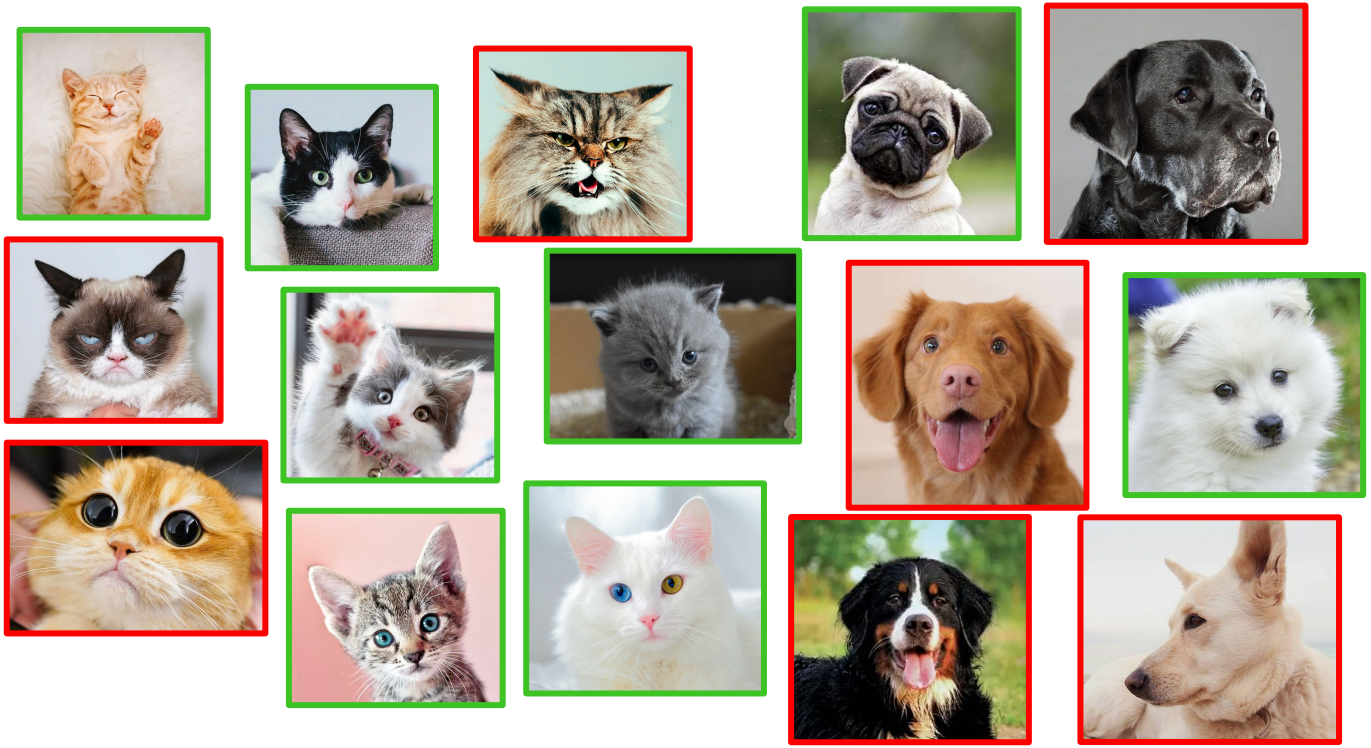
Images from various sources on the web

Performance can be assessed by analyzing true positives and negatives



Images from various sources on the web

Confusion matrix

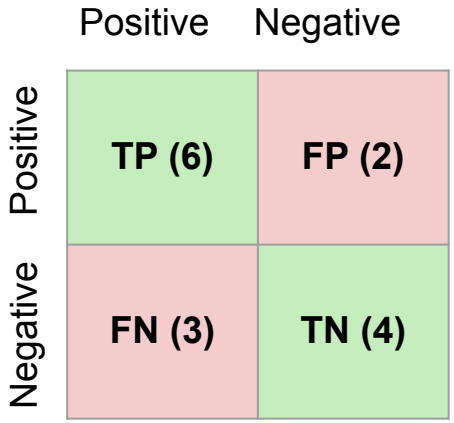


	Positive	Negative
Positive	**TP (6)**	**FP (2)**
Negative	**FN (3)**	**TN (4)**

Images from various sources on the web

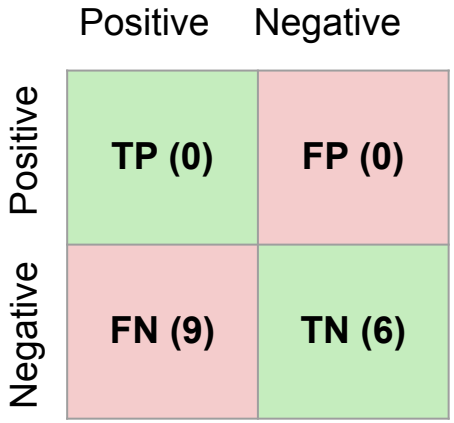
Precision tells us how well our algorithm discriminates between positives and negatives

$$\text{Precision} = \frac{TP}{TP + FP}$$



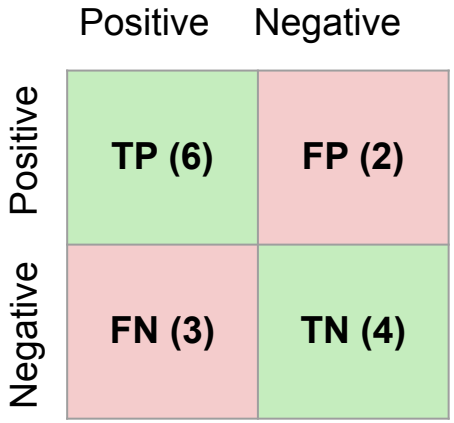
Precision tells us how well our algorithm discriminates between positives and negatives

$$\text{Precision} = \frac{TP}{TP + FP}$$



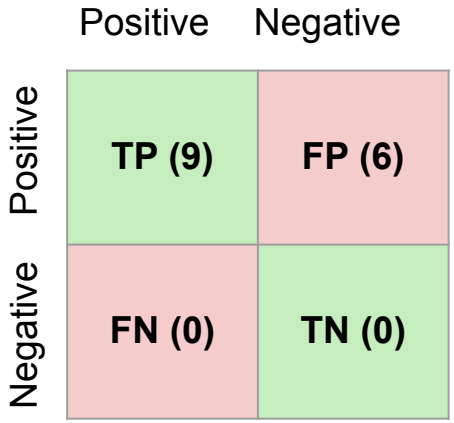
Recall tells us how good our algorithm is at finding the positives

$$\text{Recall} = \frac{TP}{TP + FN}$$



Recall tells us how good our algorithm is at finding the positives

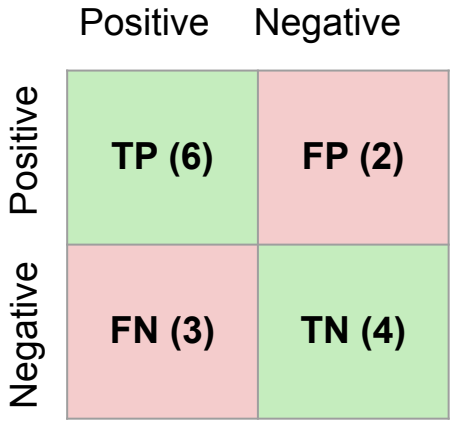
$$\text{Recall} = \frac{TP}{TP + FN}$$



It is more accurate to consider both precision and recall in our calculations

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

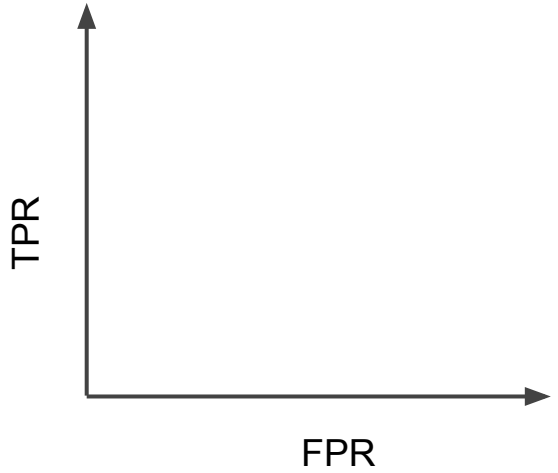


ROC curve is regularly used for evaluating performance of learning modules

ROC = Receiver Operating Characteristic curve

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

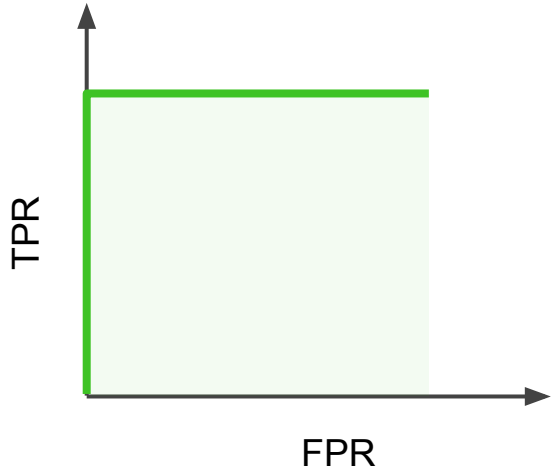


ROC curve is regularly used for evaluating performance of learning modules

ROC = Receiver Operating Characteristic curve

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

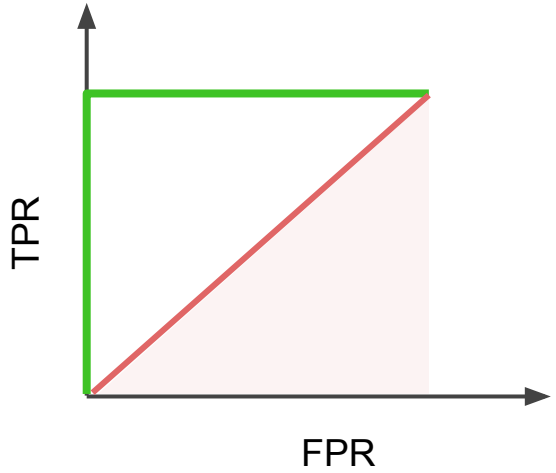


ROC curve is regularly used for evaluating performance of learning modules

ROC = Receiver Operating Characteristic curve

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

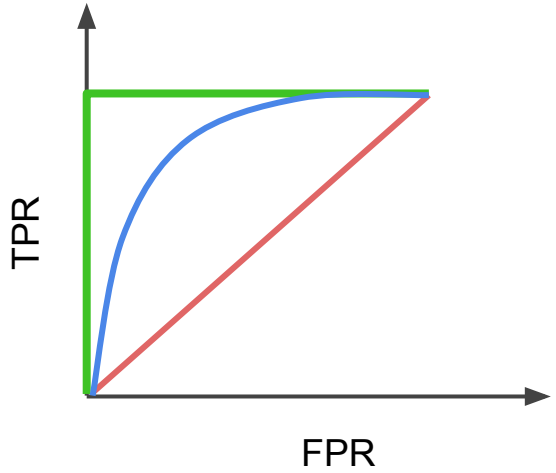


ROC curve is regularly used for evaluating performance of learning modules

ROC = Receiver Operating Characteristic curve

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$



Next lecture:

Garbage in, garbage out – the importance of input data

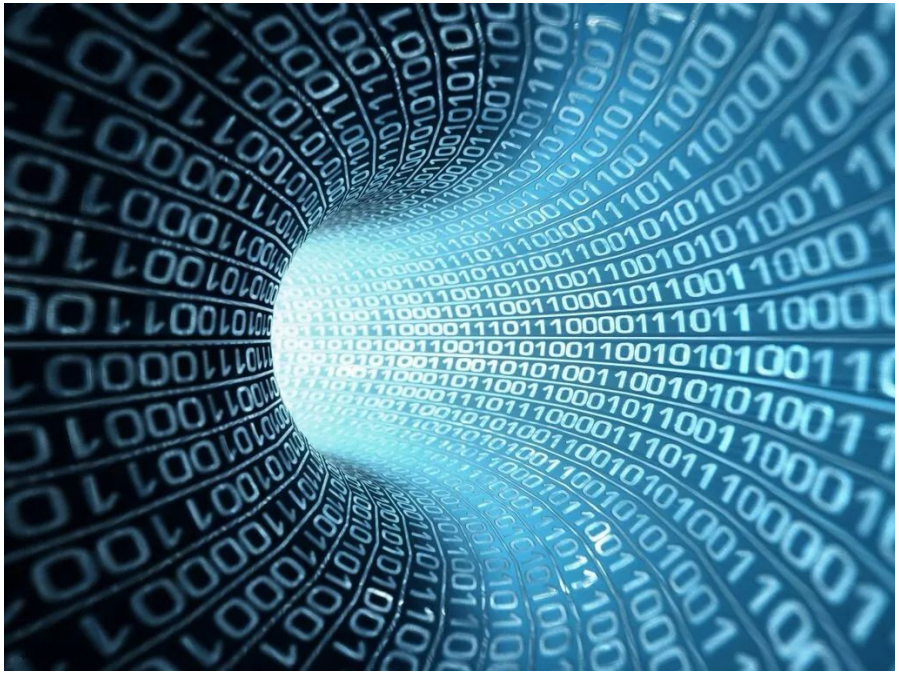


Image from GreenBook