PMILab
Pattern Recognition & Machine Intelligence

Hankuk University of Foreign Studies

A Tour of Machine Learning
A Comprehensive Overview

Prof. Jae Young Choi

Pattern Recognition and Machine Intelligence Lab. (PMI)

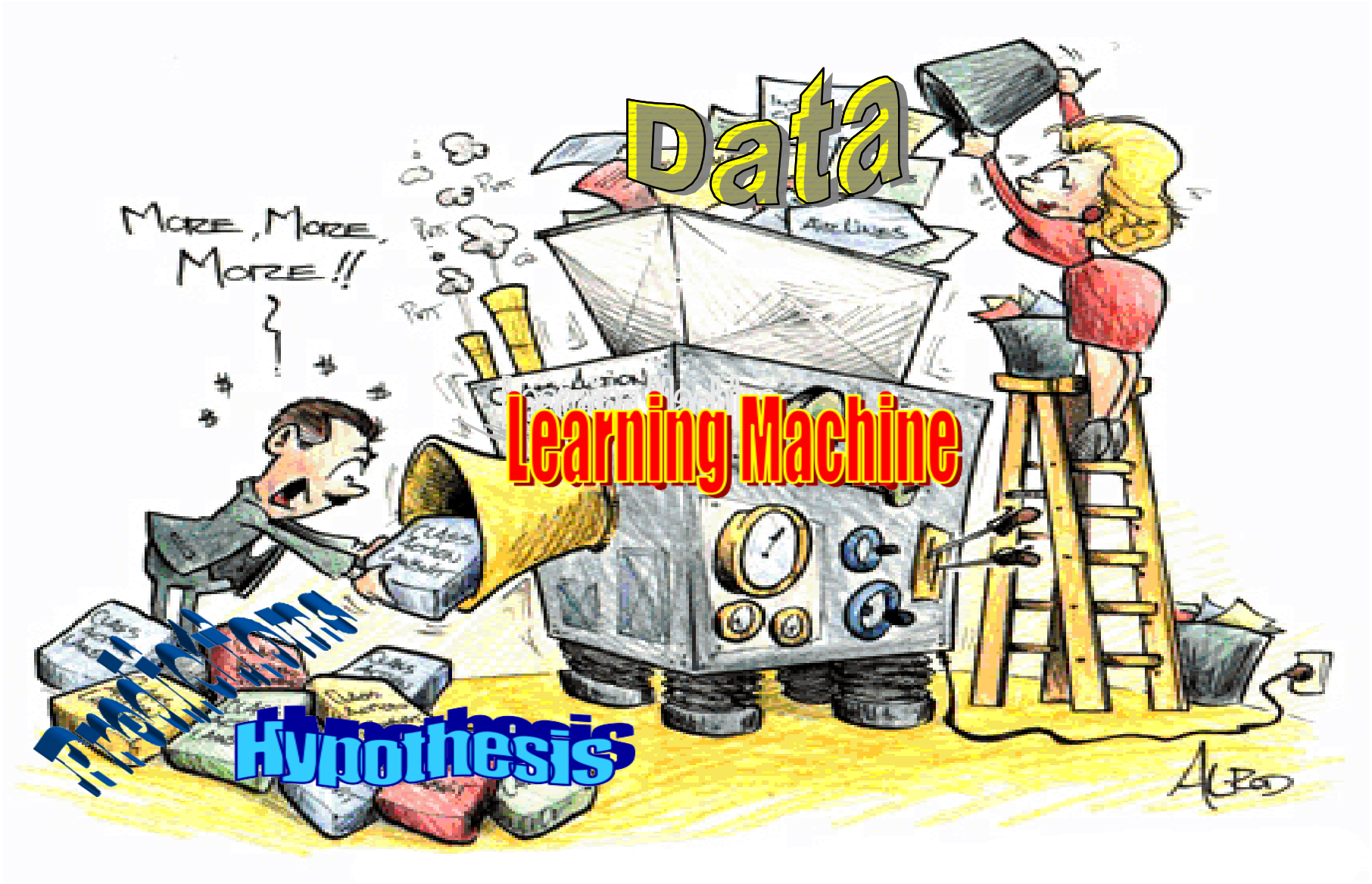
Hankuk University of Foreign Studies

WHAT IS MACHINE LEARNING?

Greek alphabet list

Upper Case Letter	Lower Case Letter	Greek Letter Name	English Equivalent
Α	α	Alpha	a
Β	β	Beta	b
Γ	γ	Gamma	g
Δ	δ	Delta	d
Ε	ε	Epsilon	e
Ζ	ζ	Zeta	z
Η	η	Eta	h
Θ	θ	Theta	th
Ι	ι	Iota	i
Κ	κ	Kappa	k
Λ	λ	Lambda	l
Μ	μ	Mu	m
Ν	ν	Nu	n
Ξ	ξ	Xi	x
Ο	ο	Omicron	o
Π	π	Pi	p
Ρ	ρ	Rho	r
Σ	σ,ς*	Sigma	s
Τ	τ	Tau	t
Υ	υ	Upsilon	u
Φ	φ	Phi	ph
Χ	χ	Chi	ch
Ψ	ψ	Psi	ps
Ω	ω	Omega	o

WHAT IS MACHINE LEARNING?



WHAT IS MACHINE LEARNING?

- **A computer program is said to learn from experience *E* with respect to tas k *T* and performance measure *P*, if its performance at tasks *T*, improves wi th experience *E***

  [Mitchell, 1997]

- **Machine learning is the systematic study of algorithms and systems that i mprove their knowledge or performance with experience**

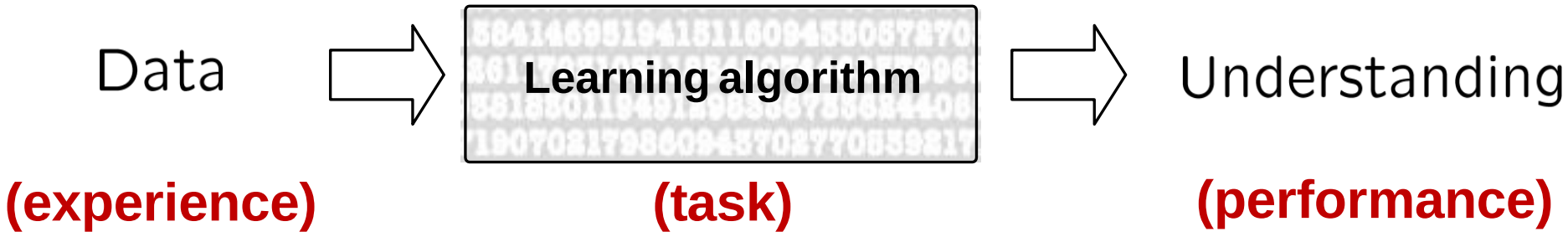
  [Peter A. Flach, 2015]

- 특정 작업 ***T*** 에서 획득한 경험적인 데이터 ***D***를 바탕으로 모델 ***M***을 자동으로 구성하여 스스로 성능 ***P***를 향상하는 컴퓨터 프로그램

WHAT IS MACHINE LEARNING?

- Study of algorithms that
  - improve their performance
  - at some task
  - with experience

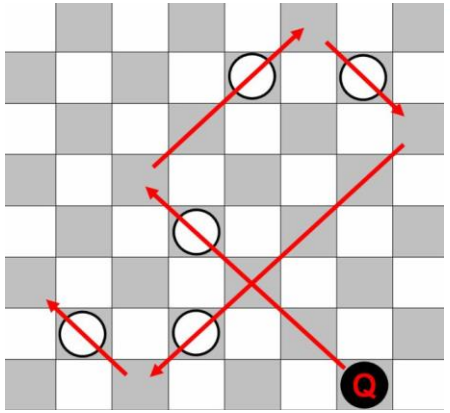
well-defined learning task: <P,T,E>



WHAT IS MACHINE LEARNING?

A checkers Learning Problem:

- Task ***T*** : playing checkers
- Performance measure ***P*** : percent of games won against opponents
- Training experience ***E***: playing practice games against itself



A Handwriting Recognition Learning Problem:

- Task ***T*** : recognizing and classifying handwritten words within images
- Performance measure ***P*** : percent of words correctly classified
- Training experience ***E*** : a database of handwritten words with given classifications

MACHINE LEARNING COMPONENT

- Tens of thousands of machine learning algorithms

- But, there are common things! Every machine learning algorithm has three components:
  - **Model**
  - **Evaluation**
  - **Optimization**

hypothesis 와 처리공간

머신 러닝 특성

- 특성 1: Self-improving Systems (인공지능)
- 특성 2: Knowledge Discovery (데이터마이닝)

- 특성 3: Data-Driven SW Design (SW공학)
- 특성 4: Automatic Programming (컴퓨터공학)

WHY MACHINE LEARNING?

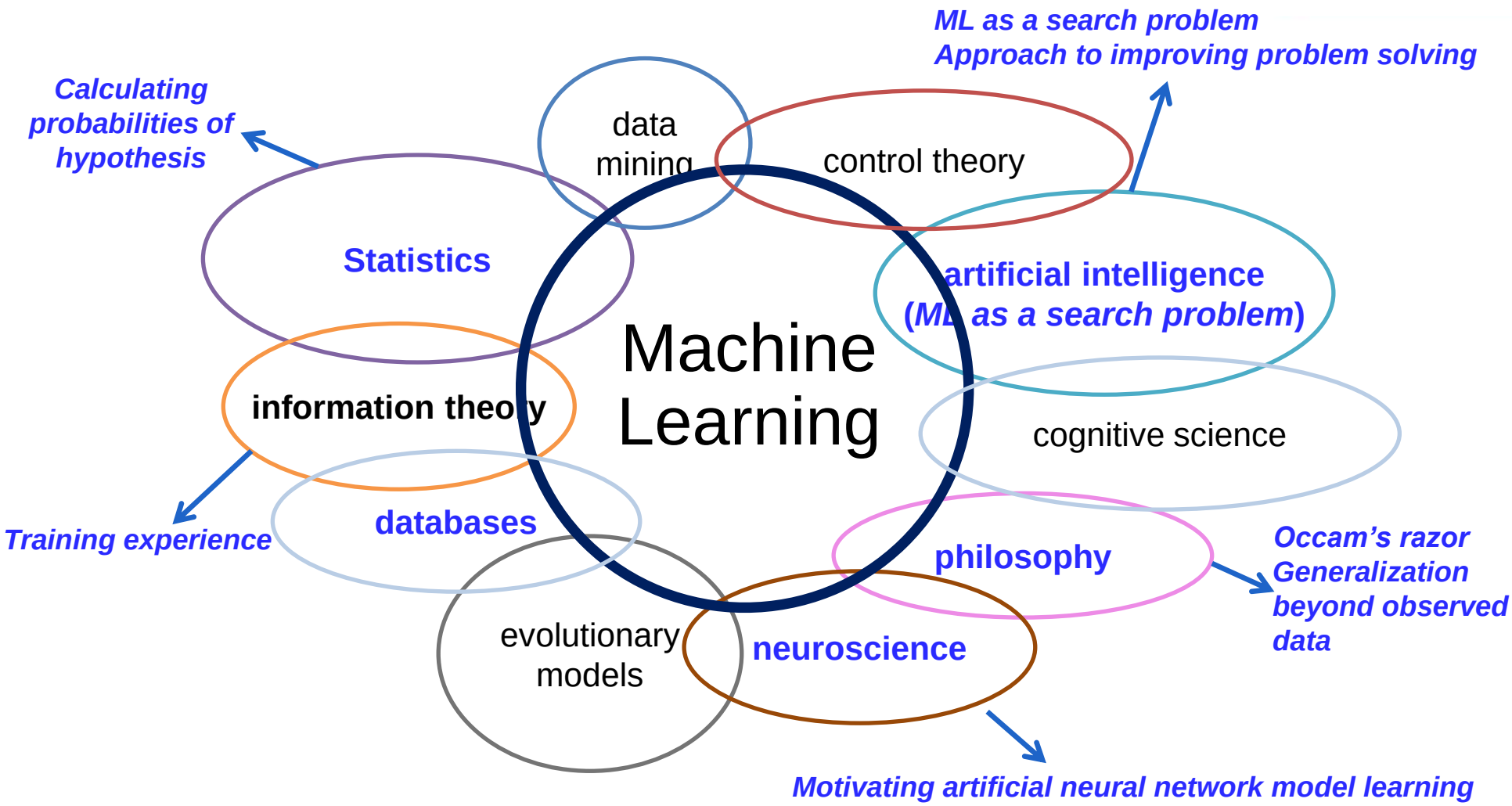
- 전문가 부족
  - industrial/manufacturing control
  - mass spectrometer analysis, drug design, astronomic discovery

- 단순 프로그래밍이 어려운 문제들
  - face/handwriting/speech recognition
  - driving a car, flying a plane

- 급격한 기술 변화
  - credit scoring, financial modeling
  - diagnosis, fraud detection

- 개인화 필요성
  - personalized news reader
  - movie/book recommendation

RELATED FIELDS



HISTORY OF MACHINE LEARNING

- **1948: Checkers Player (Samuel)**
- **1958**: **Perceptron** (Rosenblatt)
- 1975: Near miss concept learning (Winston)
- **1980: First ICML Conference (Machine Learning Workshop)**
- 1982: Self-organizing maps (Kohonen)
- 1983: Boltzmann machine (Hinton & Sejnowski)
- **1984**: **PAC computational learning theory** (Valiant)
- **1986: Backpropagation algorithm (Rumelhart, Hinton, & Williams)**
- **1986**: **Decision trees** (Quinlan)
- **1986: Machine Learning Journal**
- **1987: First NIPS Conference (Neural Information Processing Systems)**
- 1992: TD-Gammon (Tesauro)
- **1992: Support vector machines (Boser, Guyon, & Vapnik)**
- 1994: Learning Bayesian networks (Hackerman)
- 1995: Statistical learning theory (Vapnik)
- 1995: Neural Networks for Pattern Recognition (Bishop)

Checkers Player (Samuel)

ICML

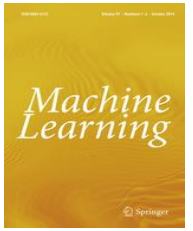
ICML Conference

Perceptron (Rosenblatt)

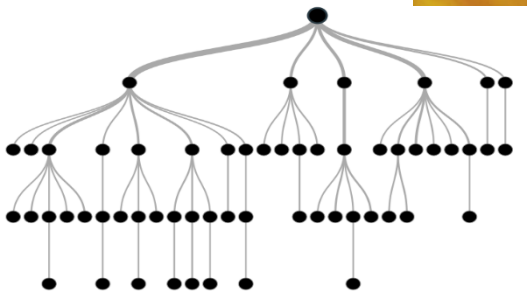
PAC Computational Learning theory (Valiant)

N I P S

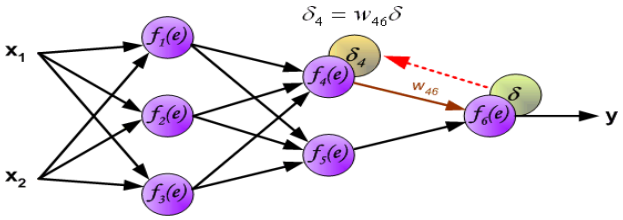
NIPS Conference



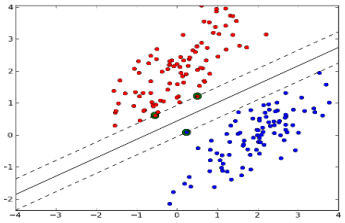
Machine Learning Journal



Decision trees



Backpropagation algorithm

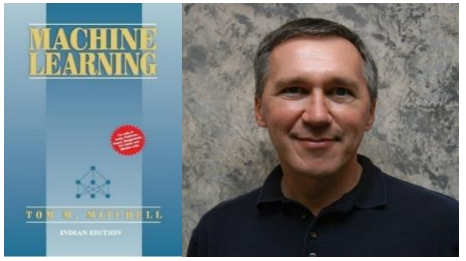


Support Vector Machine

11

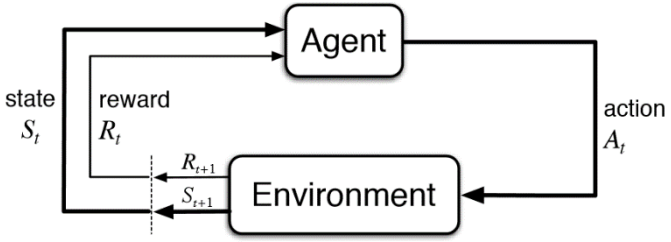
HISTORY OF MACHINE LEARNING

- **1997: Machine Learning Textbook (Mitchell)**
- 1998: Neural networks (Haykin)
- **1998**: **Reinforcement learning** (Sutton & Barto)
- 1999: Learning in graphical models (Jordan)
- 1999: Kernel machines (Schoelkopf & Smolar)
- **2001: Journal of Machine Learning Research (JMLR)**
- **2003**: **Boosting algorithms** (Freund)
- 2003: Information theory, inference, and learning algorithms (MacKay)
- 2005: Probabilistic Robotics (Thrun, Burgard, & Fox)

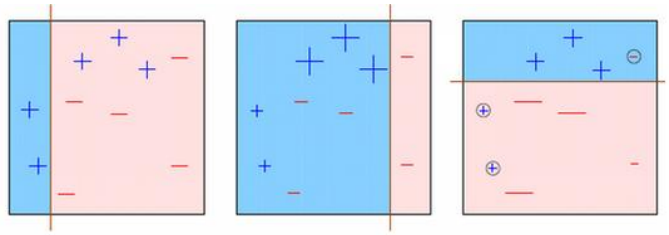


Machine Learning Textbook(Mitchell)

Journal of Machine Learning Research (JMLR)



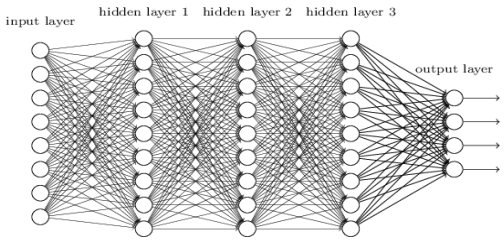
Reinforcement learning



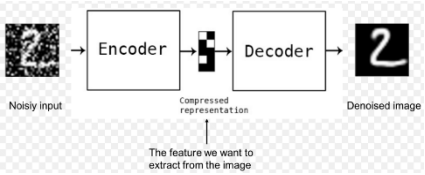
Boosting Algorithms

HISTORY OF MACHINE LEARNING

- **2006: Deep Neural Networks (DBNs) with Autoencoder (Hinton)**
- 2006: Pattern Recognition and Machine Learning (Bishop)
- 2008: Neural Networks and Learning Machines (Haykin)
- 2009: Probabilistic Graphical Models (Koller)
- 2009: Siri personal assistant (Apple)
- 2009: Google Car (Thrun)
- **2011**: **Watson AI supercomputer (IBM)**
- **2012: AlexNet (DCNN) wins ImageNet** (Alex Krizhevsky)
- 2012: DNNresearch deep learning (Hinton & Google)
- 2012: Large-scale image retrieval (Google)
- 2013: Human Brain Project HBP (EU)
- **2013: Quantum AI Lab for machine learning (Google)**
- **2014: Generative Adversarial Networks (Goodfellow)**
- 2015: Institute of Deep Learning (Baidu)

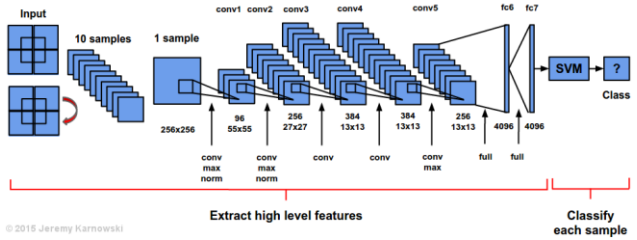


Deep Neural Netw

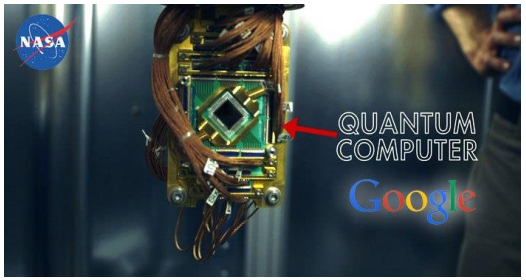


Autoencoder

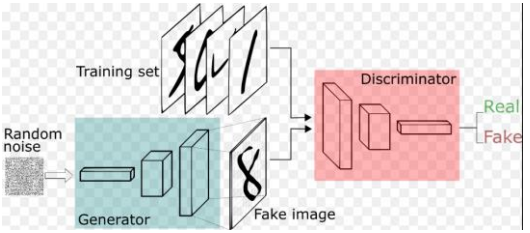
Watson AI supercomputer



AlexNet



Quantum AI Lab



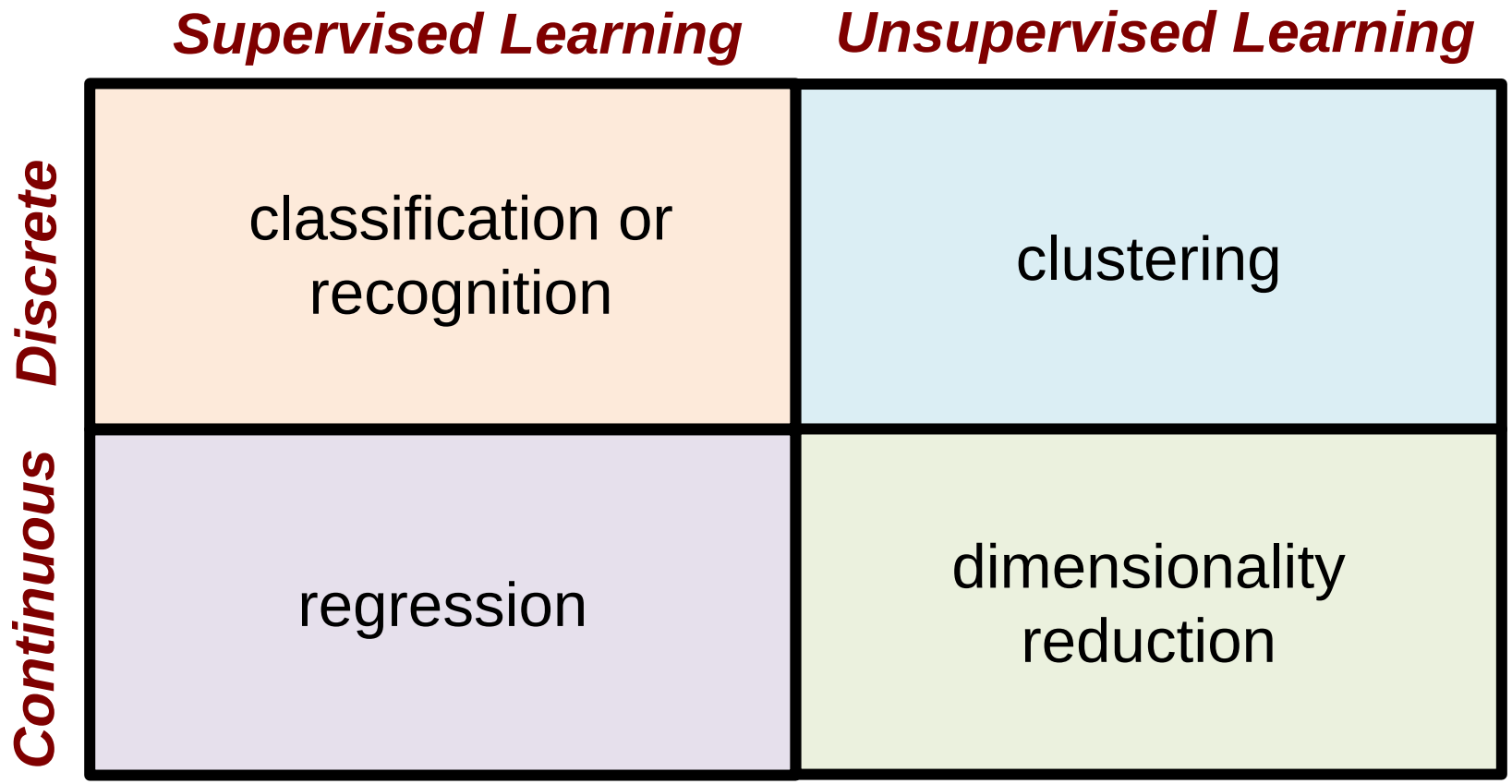
GANs

TYPES OF MACHINE LEARNING APPROACH

THREE DIFFERENT TYPES

- **Supervised (inductive) Learning**
  - Training data includes desired outputs
- **Unsupervised Learning**
  - Training data does not include desired outputs
- **Reinforcement Learning**
  - Rewards from sequence of actions

TYPES OF MACHINE LEARNING



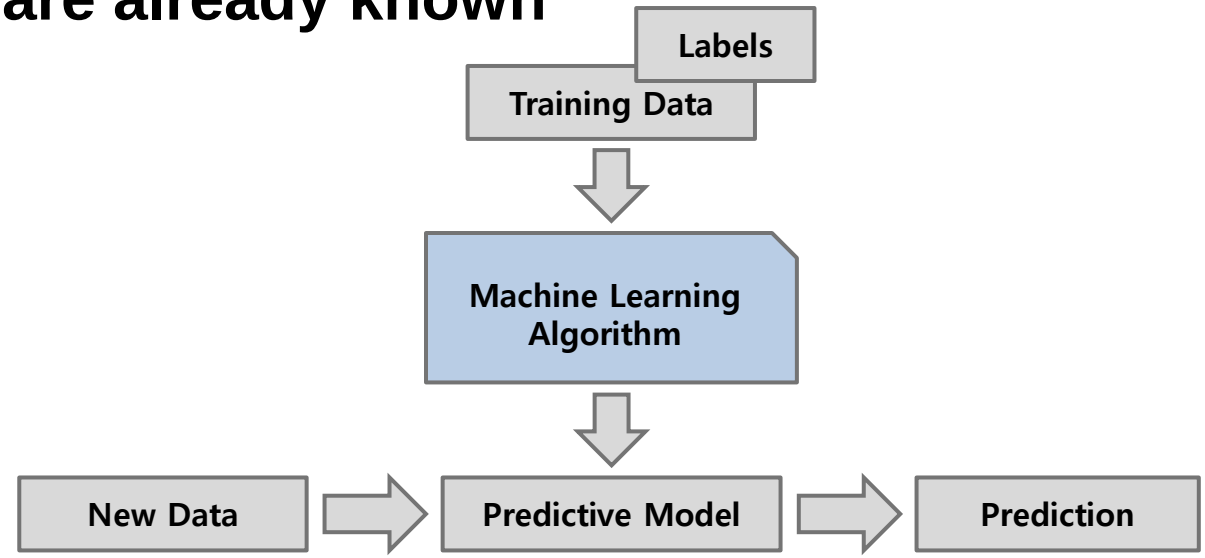
머신러닝 학습 방법 및 모델 구조

학습 방법	학습 문제의 예
감독 학습	**인식, 분류, 진단, 예측, 회귀분석**
무감독 학습	**군집화, 밀도추정, 차원 축소, 특징추출**
강화 학습	**시행착오, 보상 함수, 동적 프로그래밍**

모델 구조	표현	기계학습 모델 예
논리식	명제 논리, 술어논리, Prolog 프로그램	Version Space, 귀납적 논리 프로그래밍(ILP)
규칙	If-Then 규칙, 결정규칙	AQ
함수	Sigmoid, 다항식, 커널	신경망, RBF망, SVM, 커널머신
트리	유전자 프로그램, Lisp 프로그램	결정 트리, 유전자 프로그래밍, 뉴럴트리
그래프	방향성/무방향성 그래프, 네트워크	확률그래프 모델, 베이지안망, HMM

SUPERVISED LEARNING

- **Learn a model from labeled training data**
- **Make classification/predictions about new (unseen) data**
- **“Supervised” refers to a set of samples where the desired output signals (labels) are already known**
- **Two categories**
  - Classification task
  - Regression task



SUPERVISED LEARNING (FORMAL DEFINITION)

- **Given examples of a function *(X, F(X))***

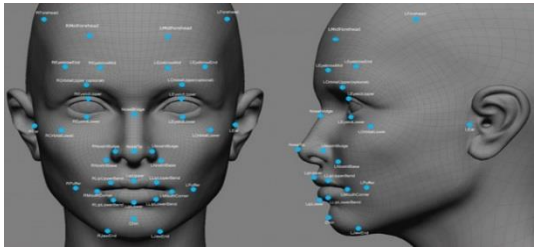
- **Predict function *F(X)* for new examples *X***
  - Discrete *F(X)*: Classification
  - Continuous *F(X)*: Regression
  - *F(X)* = Probability(*X*): Probability estimation

Example Applications

- **Credit risk assessment**
  x: Properties of customer and proposed purchase.
  f(x): Approve purchase or not.

- **Disease diagnosis**
  x: Properties of patient (symptoms, lab tests)
  f(x): Disease (or maybe, recommended therapy)

- **Face recognition**
  x: Bitmap picture of person's face
  f(x): Name of the person.



- **Automatic Steering**
  x: Bitmap Picture of road surface in front of car.
  f(x): Degrees to turn the steering wheel.

SUPERVISED LEARNING-CLASSIFICATION

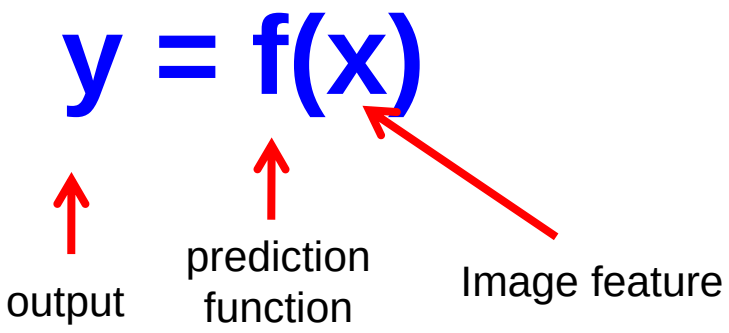
- **Apply a prediction function to a feature representation of the image to get the desired output:**

f() = “apple”

f() = “tomato”

f() = “cow”

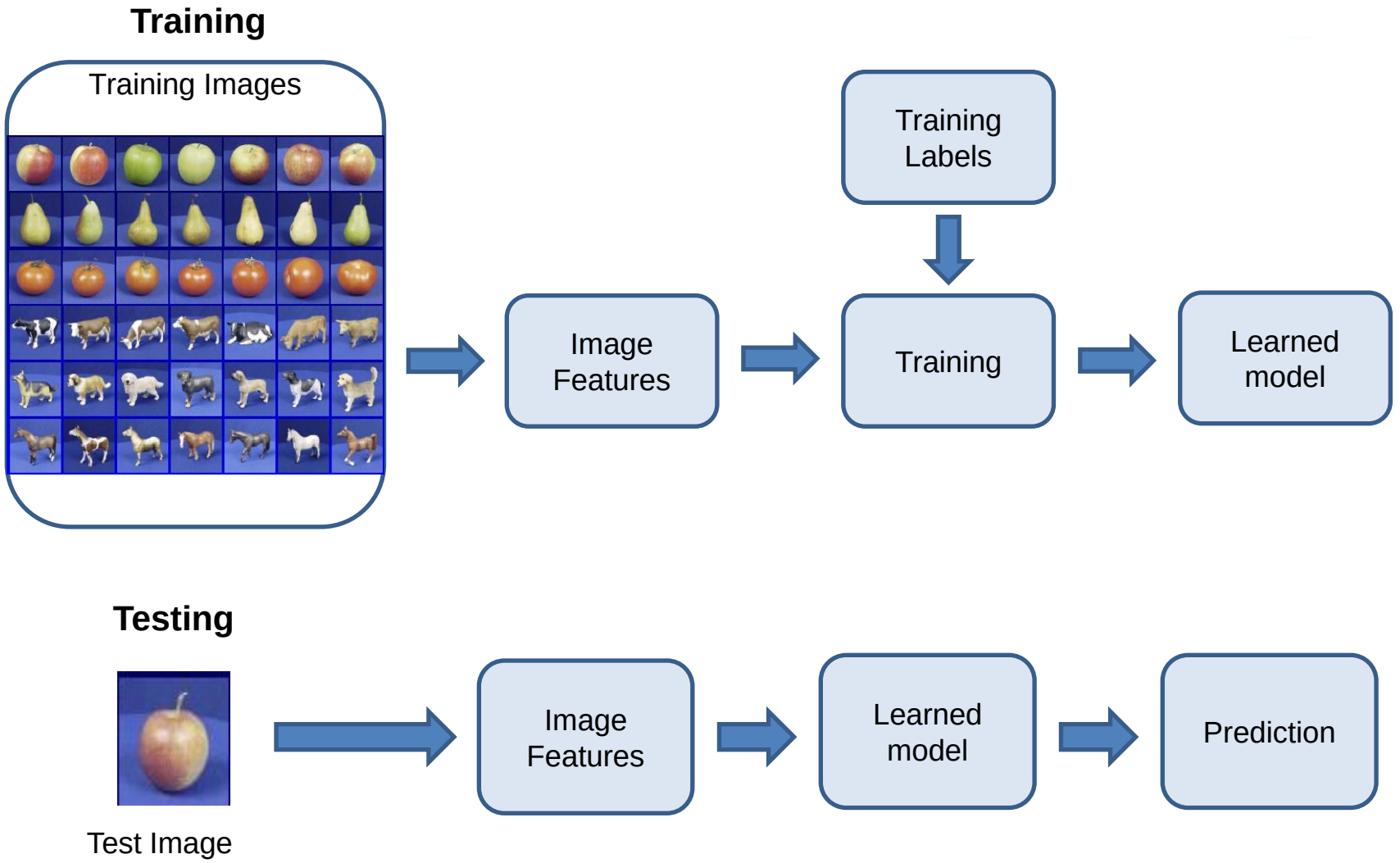
SUPERVISED LEARNING-CLASSIFICATION



- **Training:** given a *training set* of labeled examples $\{(x_1,y_1), \ldots, (x_N,y_N)\}$, estimate the prediction function $f(\cdot)$ by minimizing the prediction error on the training set

- **Testing:** apply $f(\cdot)$ to a never before seen *test example* x and output the predicted value $y = f(x)$

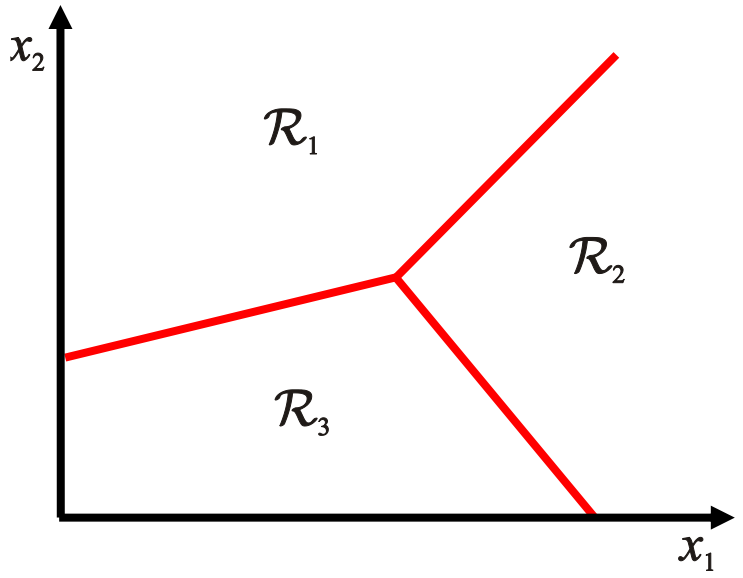
SUPERVISED LEARNING-CLASSIFICATION



SUPERVISED LEARNING-CLASSIFICATION FROM GEOMETRIC PERSPECTIVE

- **Assign input vector to one of two or more classes**

- **Any decision rule divides input space into *decision regions* separated by *decision boundaries***

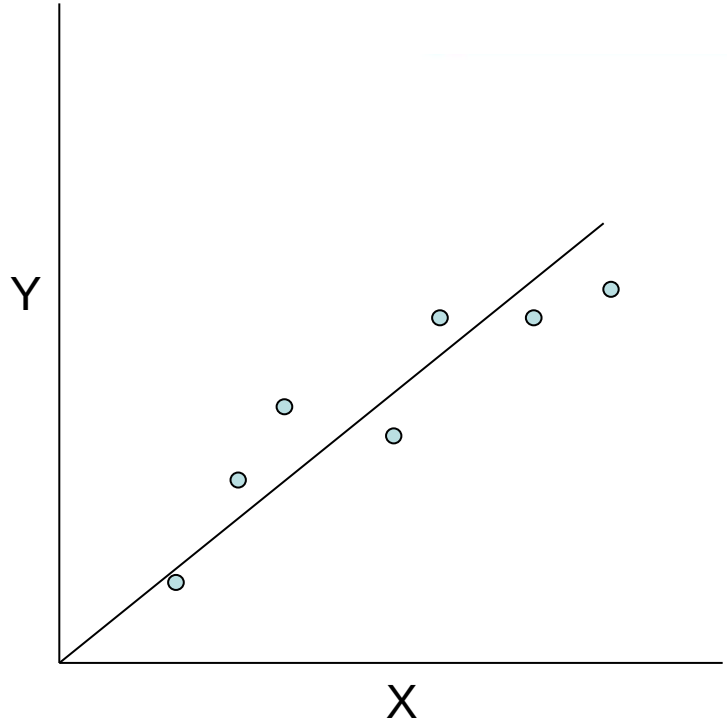


SUPERVISED LEARNING-CLASSIFICATION (TYPES OF SOME CLASSIFIERS)

- **K-Nearest Neighbor (K-NN)**
- **Support Vector Machines (SVMs)**
- **Boosted Decision Trees**
- **Neural networks (NNs)**
- **Naïve Bayes**
- **Logistic regression**
- **Randomized Forests**
- **Etc.**

SUPERVISED LEARNING - REGRESSION

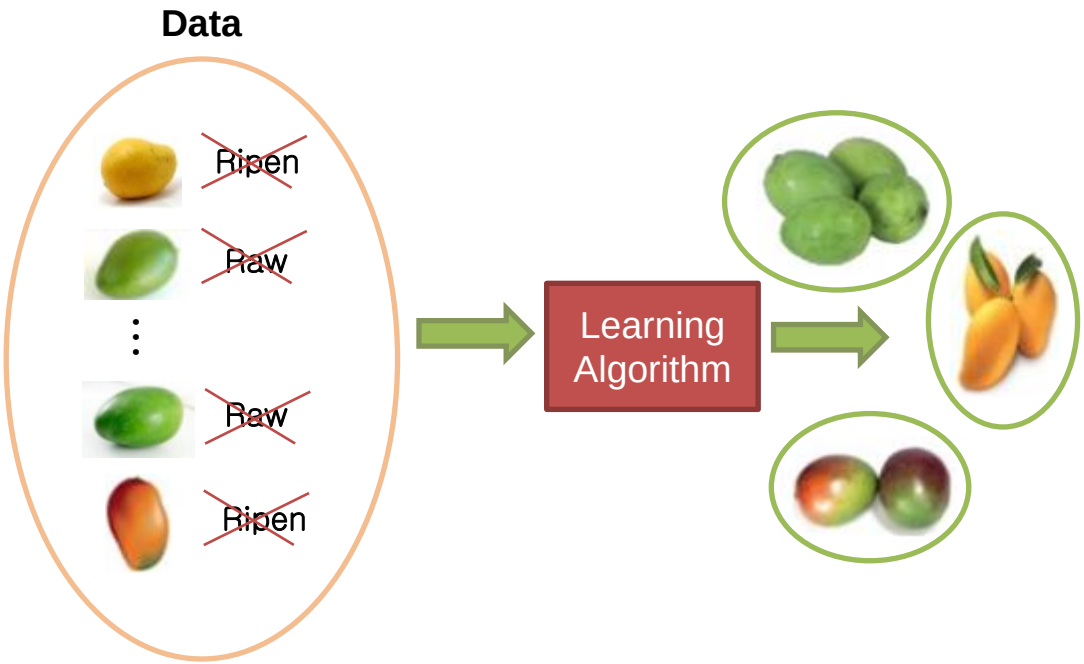
- **Given an input x we would like to compute an output y**
  - Given a number of predictor (explanatory) variables
  - Outcome (response variable)
- **For example:**
  - Predict height from age
  - Predict Google's price from Yahoo's price
  - Predict distance from wall using sensor readings



Note that now Y can be **continuous**

UNSUPERVISED LEARNING

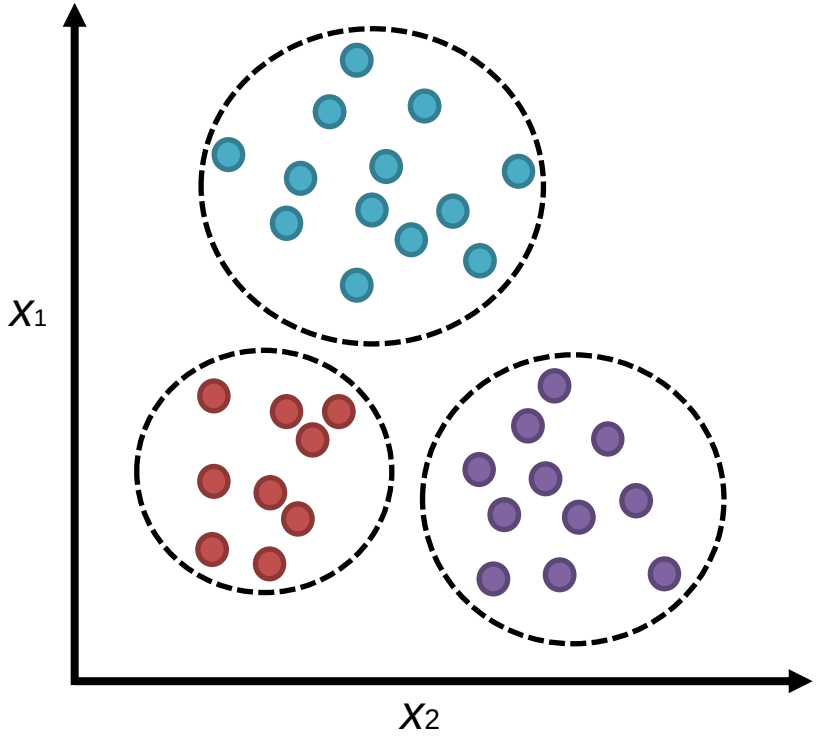
- **We don’t know the *right answer* before we train our model**
- **Dealing with *unlabeled* data or data of *unknown structure***
- **Using supervised learning, we can explore the *structure of our data* to extract meaningful information without guidance…**
  - Known outcome variable
  - Reward function



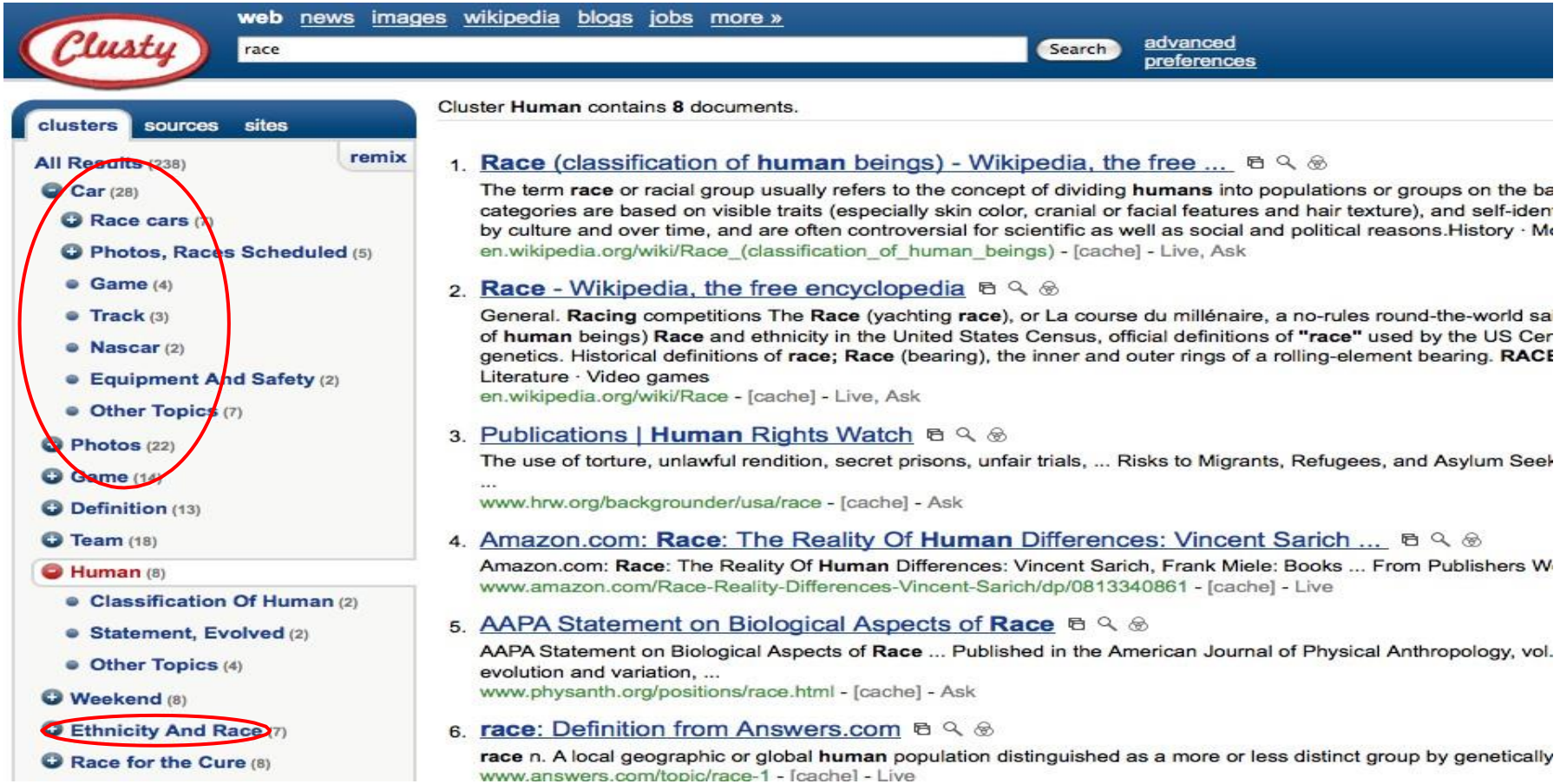
UNSUPERVISED LEARNING - CLUSTERING

- ***Clustering*** is an exploratory data analysis that organizes a pile of information into ***meaningful subgroups* (*clusters*)** without having any prior knowledge of their group memberships

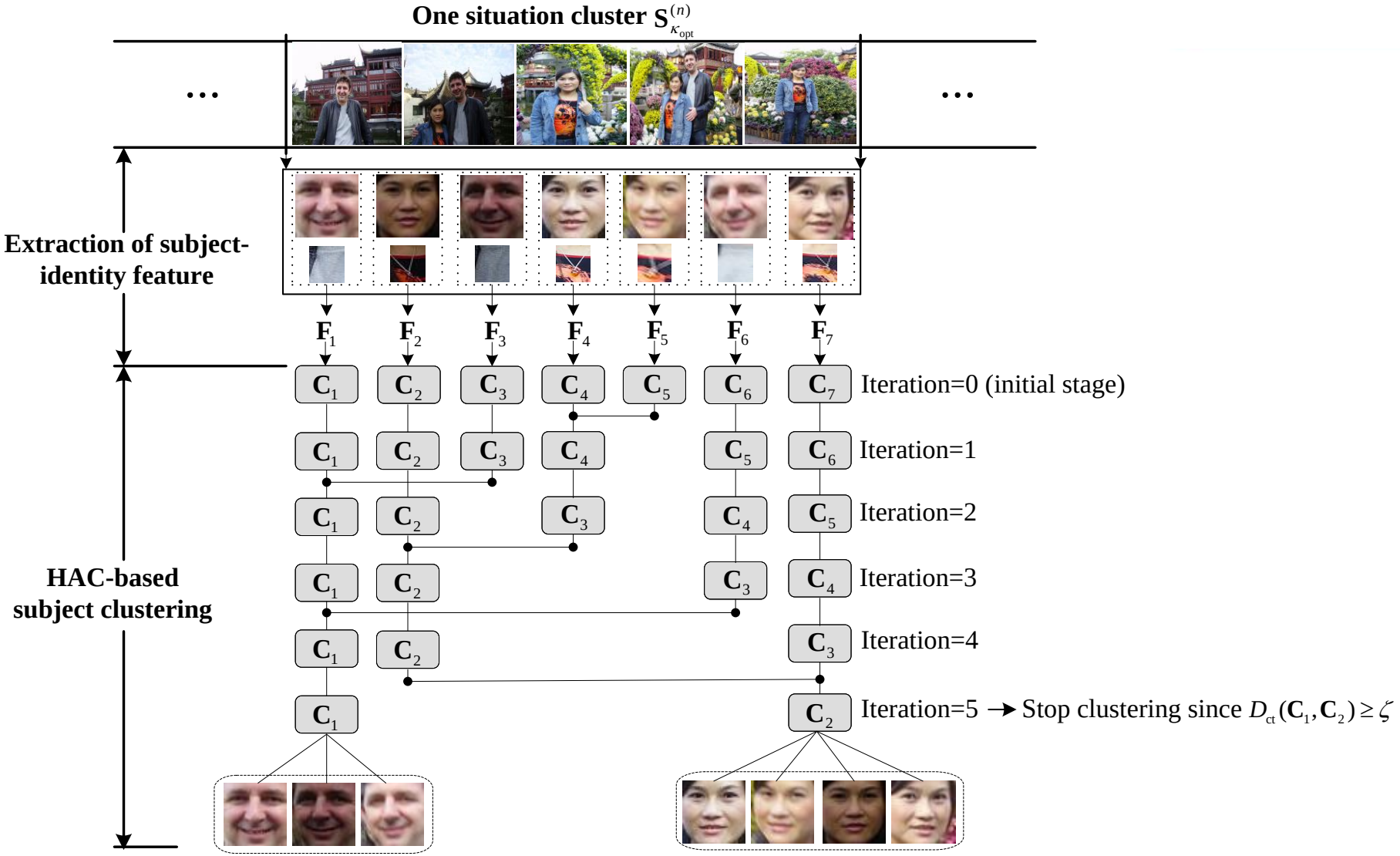
- The figure (on right-hand side) illustrates how clustering can be applied to organizing unlabeled data into three distinct groups based on similarity of their features x_1 and x_2



UNSUPERVISED LEARNING – CLUSTERING WEB SEARCH RESULTS



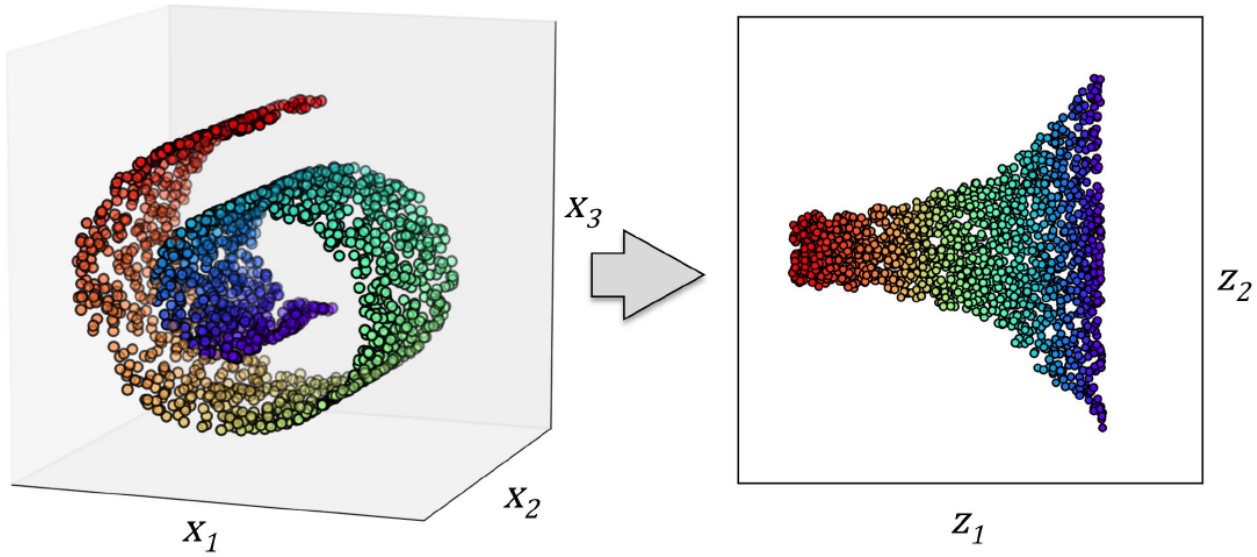
UNSUPERVISED LEARNING – CLUSTERING PHOTO CATEGORIZATION BASED ON PERSON



UNSUPERVISED LEARNING-DIMENSIONALITY REDUCTION

- **For feature processing to remove data from noise**

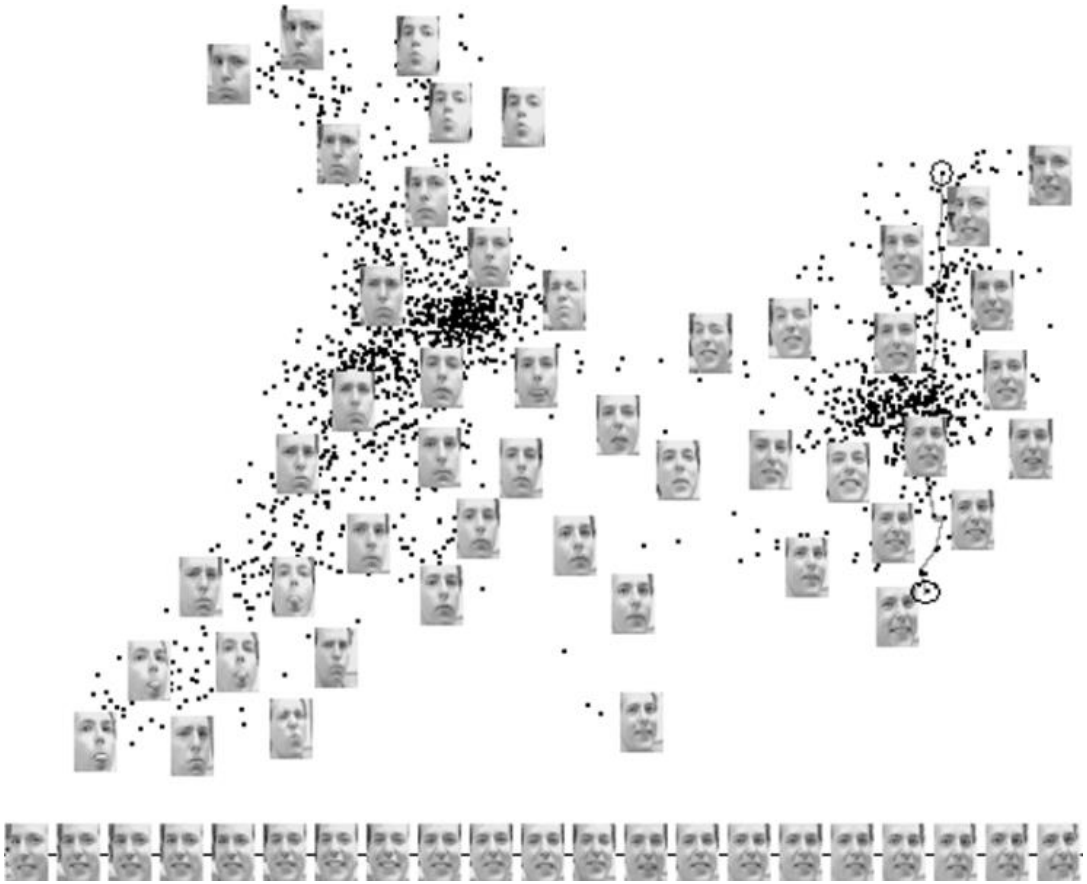
- **Compress the data onto a smaller dimensional subspace while retaining most of the relevant information**



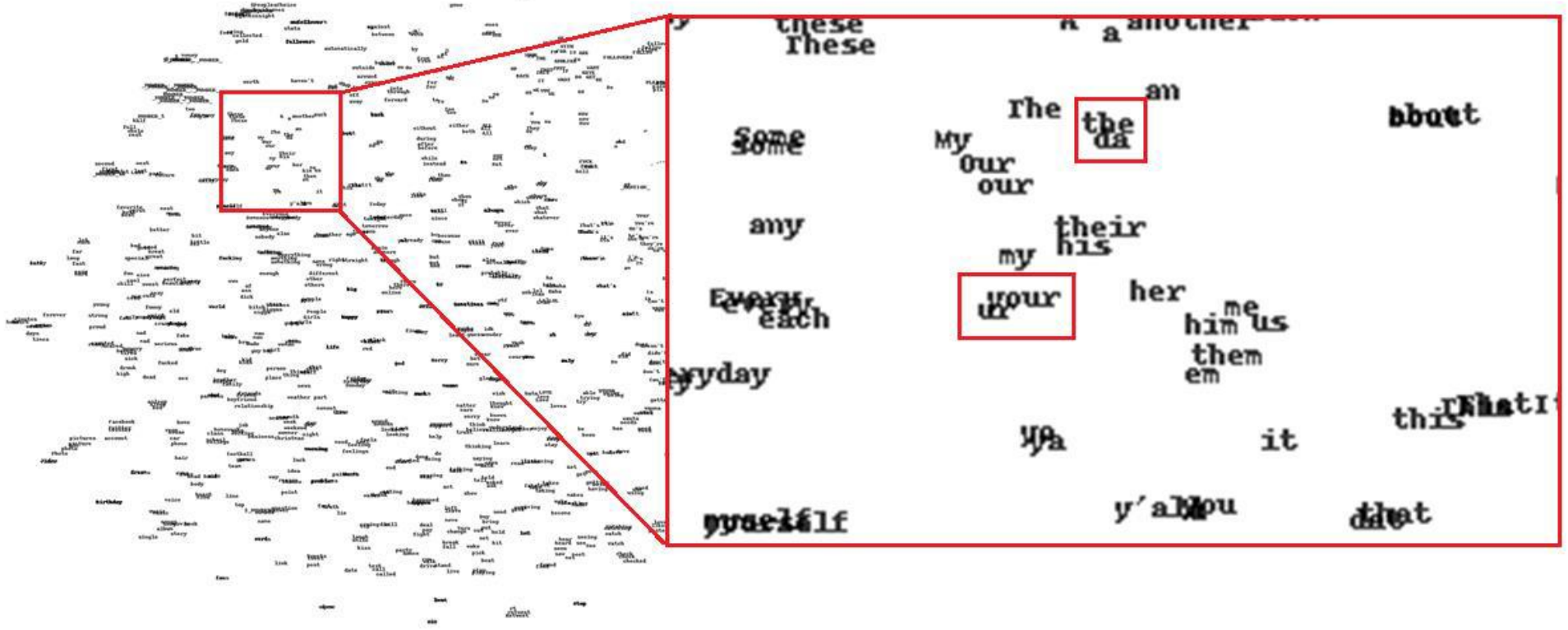
UNSUPERVISED LEARNING-DIMENSIONALITY REDUCTION MAINFOLD LEARNING

Images have thousands or millions of pixels

Can we map each image into subspace (manifold) such that similar images are near each other?



UNSUPERVISED LEARNING-DIMENSIONALITY REDUCTION WORD EMBEDDING FOR NLP

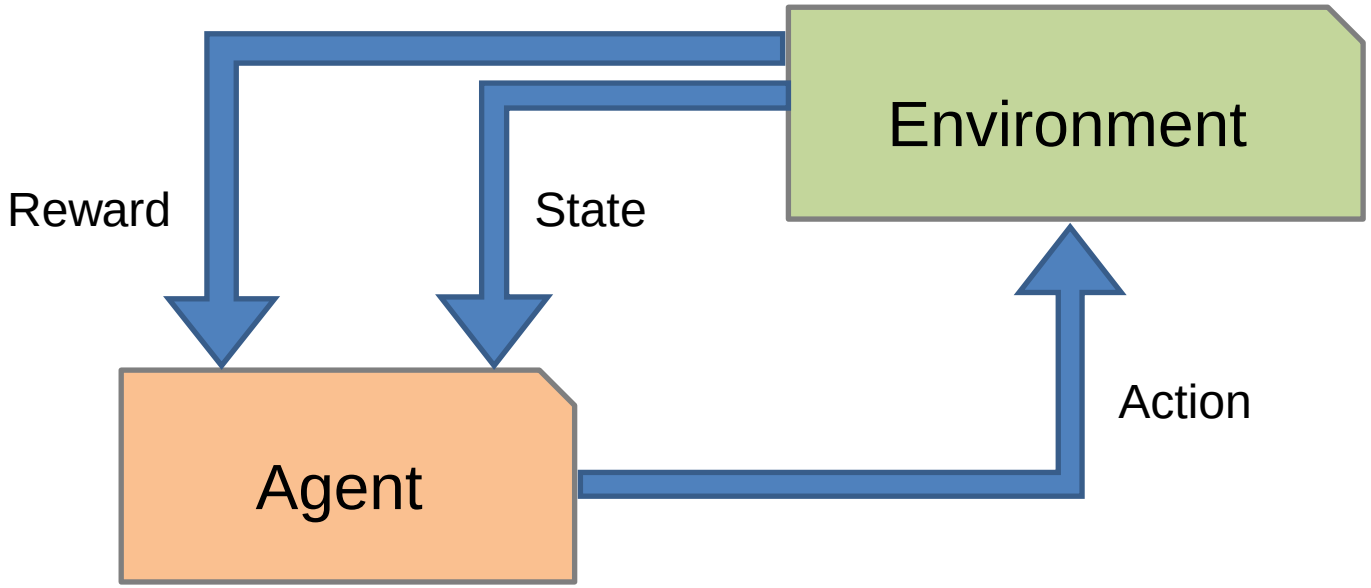


2-D projection of the top 100 words used on Twitter

REINFORCEMENT LEARNING

- Goal is to develop a system (***agent***) that improves its performance based on ***interaction with environment***
- Instead of using ground truth label, a ***reward function*** is used

  → Measuring of how well the action
- Through the interaction with the environment, an agent can learn a series of actions that ***maximizes the reward*** via an exploratory ***trial-and-error*** approach



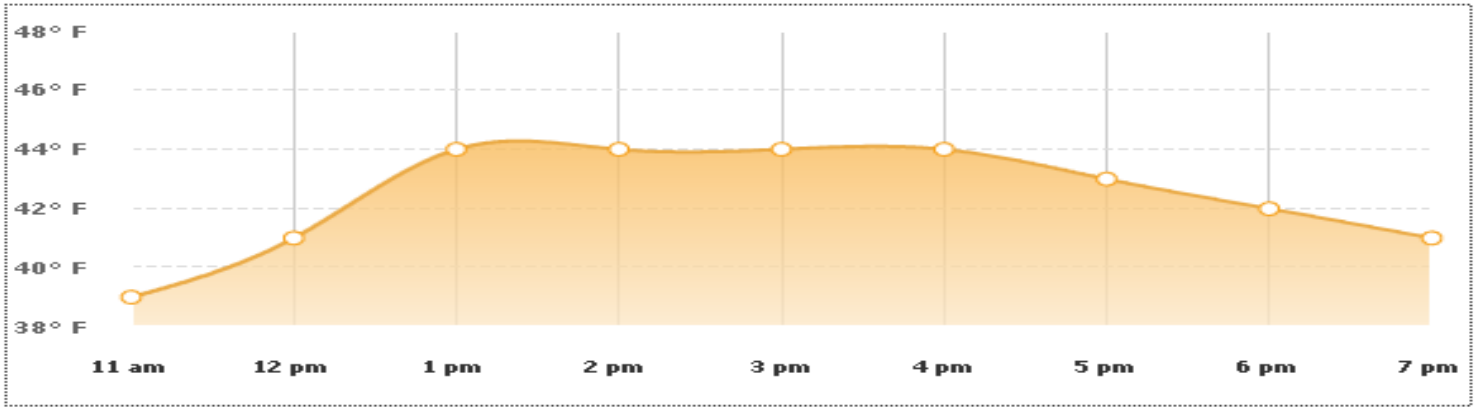
Pop Quiz 1

- **Supervised/Unsupervised?** **Labels?** **Classification/Regression?**

Supervised

Temperature

Regression



Temperature prediction for weather forecasting

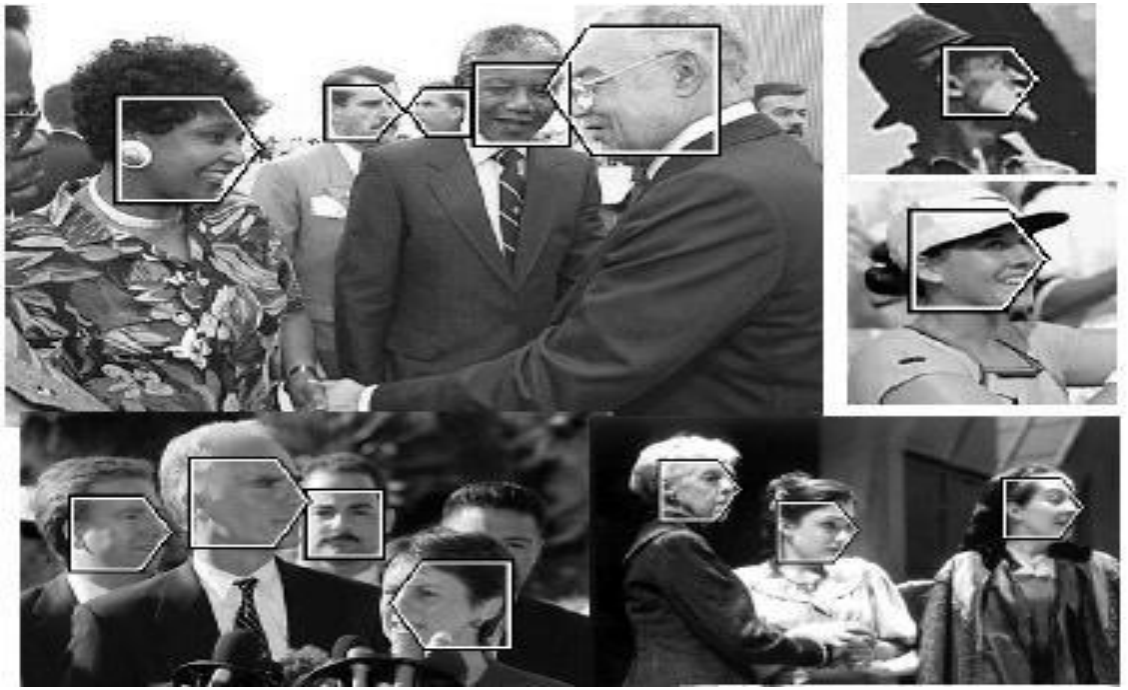
Pop Quiz 2

- **Supervised/Unsupervised?** **Labels?** **Classification/Regression?**

Supervised

Pos: face , Neg: non-face

classification



Face Detection