

# Introduction to Machine Learning

Tanujit Chakraborty

Indian Statistical Institute, Kolkata.

Email: [tanujitisi@gmail.com](mailto:tanujitisi@gmail.com)

July 10, 2019



- “**Statistics** is the universal tool of inductive inference, research in natural and social sciences, and technological applications.

**Statistics**, therefore, must always have purpose, either in the pursuit of knowledge or in the promotion of human welfare”

- P.C. Mahalanobis, Father of Statistics in India.

- **Role of Statistics:**

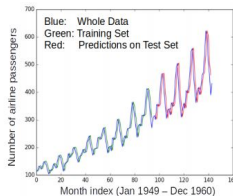
- ① making inference from samples
- ② development of new methods for complex data sets
- ③ quantification of uncertainty and variability

- **Remember:** “Figure won’t lie, but liars figure”

- “Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed”
  - Arthur L. Samuel, AI pioneer.
- Role of Machine Learning: efficient algorithms to
  - 1 solve an optimization problem
  - 2 represent and evaluate the model for inference
  - 3 create programs that can automatically learn rules from data
- Remember: “Prediction is very difficult, especially if it’s about the future” - - Niels Bohr, Father of Quantum.

# Introduction to Machine Learning

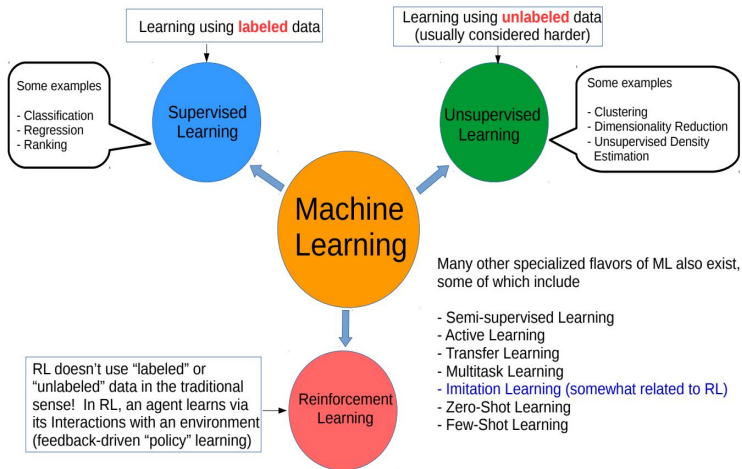
- Designing algorithms that **ingest data** and **learn a model** of the data.
- The learned model can be used to
  - ① Detect **patterns/structures/themes/trends** etc. in the data
  - ② Make **predictions** about future data and make decisions



- Modern ML algorithms are heavily **“data-driven”**.
- Optimize a performance criterion using example data or **past experience**.

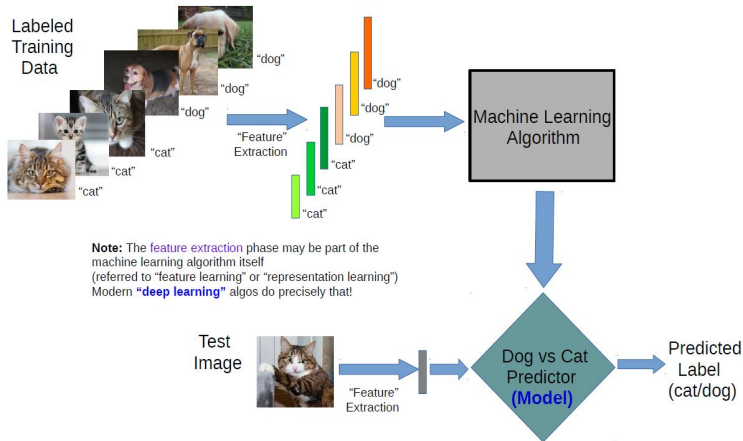
# Taxonomy for Machine Learning

Machine learning provides systems the ability to automatically learn



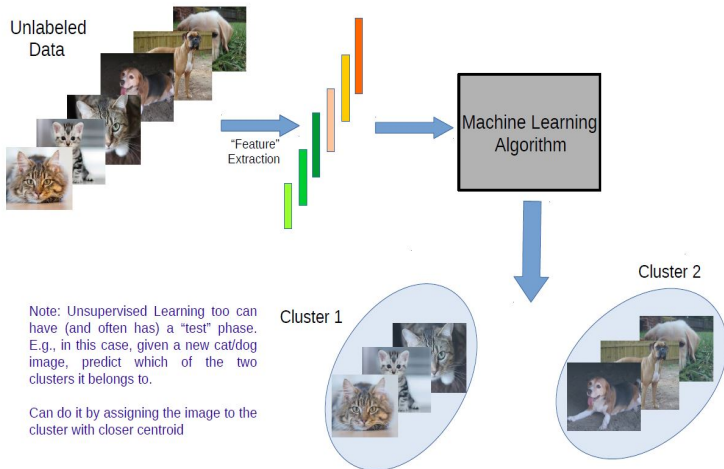
# A Typical Supervised Learning Workflow (for Classification)

**Supervised Learning:** Predicting patterns in the data



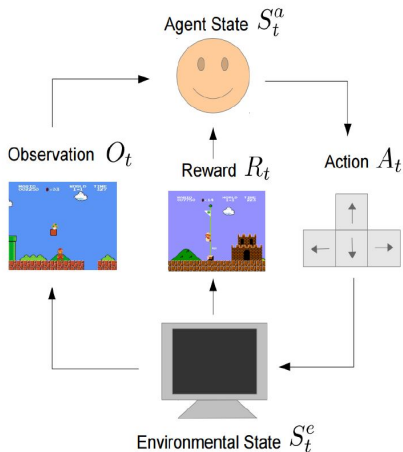
# A Typical Unsupervised Learning Workflow (for Clustering)

**Unsupervised Learning:** Discovering patterns in the data



# A Typical Reinforcement Learning Workflow

**Reinforcement Learning:** Learning a "policy" by performing actions and getting rewards (e.g, robot controls, beating games)



Agent's goal is to learn a policy for some task

Agent does the following repeatedly

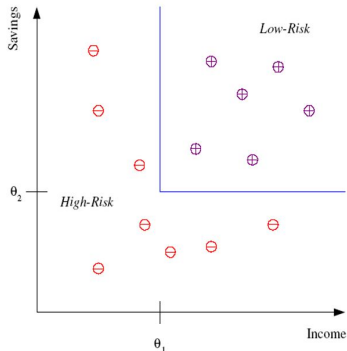
- Senses/observes the environment
- Takes an action based on its current policy
- Receives a reward for that action
- Updates its policy

There IS supervision, not explicit (as in Supervised Learning) but rather implicit (feedback based)

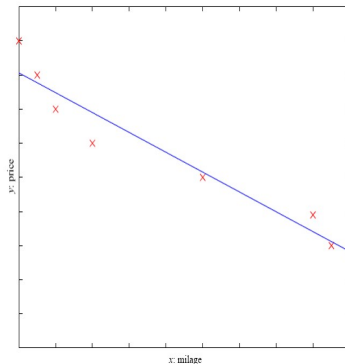


# Classification

- **Example:** Credit scoring.
- Differentiating between low-risk and high-risk customers from their income and savings.
- **Discriminant:** IF  $\text{Income} > \theta_1$  AND  $\text{Savings} > \theta_2$  THEN low-risk ELSE high-risk.
- **Classification:** Learn a linear/nonlinear separator (the “model”) using training data consisting of input-output pairs (each output is discrete-valued “label” of the corresponding input).
- Use it to predict the labels for new “test” inputs.
- **Other Applications:** Image Recognition, Spam Detection, Medical Diagnosis.

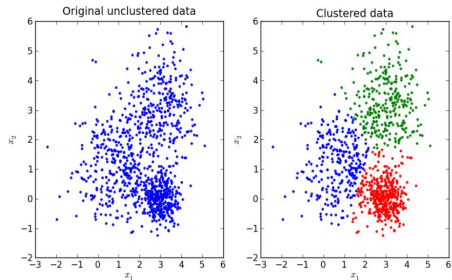


- **Example:** Price of a used car.
- $X$  : car attributes;  $Y$  : price and  $Y = f(X, \theta)$
- $f(\cdot)$  is the model and  $\theta$  is the model parameters.
- **Regression:** Learn a line/curve (the “model”) using training data consisting of Input-output pairs (each output is a real-valued number).
- Use it to predict the outputs for new “test” inputs.
- **Other Applications:** Price Estimation, Process Improvement, Weather Forecasting.



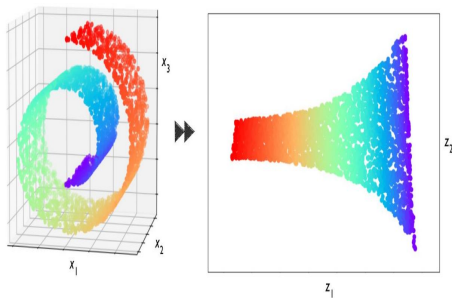
# Clustering

- **Given:** Training data in form of unlabeled instances  $\{x_1, x_2, \dots, x_N\}$
- **Goal:** Learn the intrinsic latent structure that summarizes/explains data
- **Clustering:** Learn the grouping structure for a given set of unlabeled inputs.
- Homogeneous groups as latent structure: **Clustering**
- **Other Applications:** Topic Modelling, Image Segmentation, Social Networking.



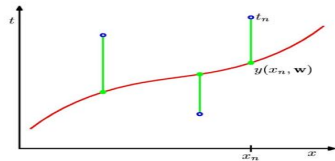
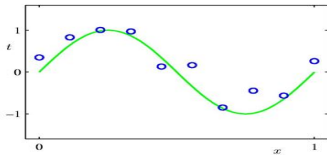
# Dimensionality Reduction

- Low-dimensional latent structure:  
Dimensionality Reduction
- **Goal:** Learn a Low-dimensional representation for a given set of high-dimensional inputs
- **Note:** DR also comes in supervised flavors (supervised DR).
- **Figure:** Three-dimension to two-dimension nonlinear projection (a.k.a. manifold learning)



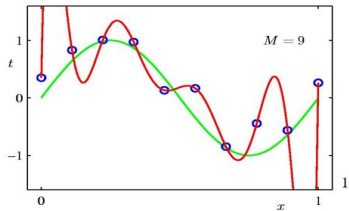
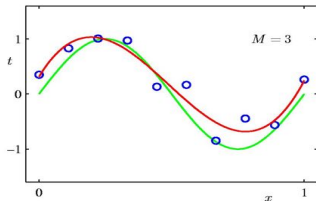
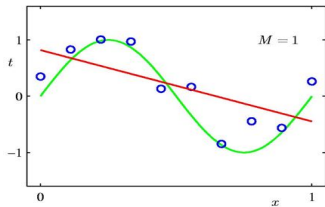
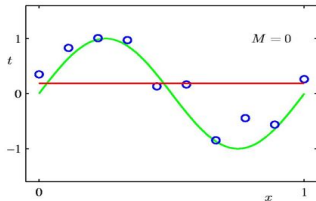
# A Simple Example: Fitting a Polynomial

- The **green curve** is the true function (which is not a polynomial).
- The data points are uniform in  $x$  but have noise in  $y$ .
- We will use a loss function that measures the **squared error** in the prediction of  $y(x)$  from  $x$ . The loss for the **red polynomial** is the sum of the squared vertical errors.



# Some fits to the data: which is best?

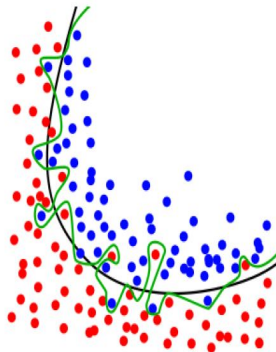
The right model complexity?



**Desired:** hypotheses that are not too simple, not too complex (so as to not overfit on the training data)

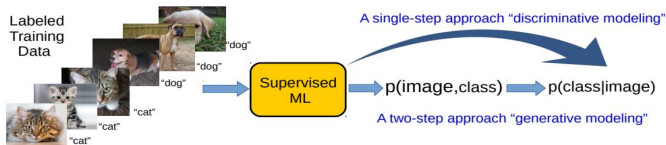
# Overfitting and Generalization

- Doing well on the training data is not enough for an ML algorithm.
- Trying to do too well (or perfectly) on training data may lead to bad “generalization”.
- **Generalization**: Ability of an ML algorithm to do well on future “test” data.
- Simple models/functions tend to prevent **overfitting** and generalize well: A key principle in designing ML algorithms (called “**regularization**”)
- **No Free Lunch Theorem**

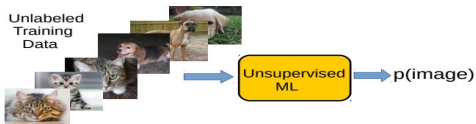


# Probabilistic Machine Learning

- **Supervised Learning** (“predict  $y$  given  $x$ ”) can be thought of as estimating  $p(Y|X)$



- **Unsupervised Learning** (“model  $x$ ”) can also be thought of as estimating  $p(x)$

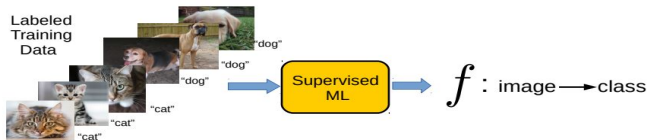


- Harder for Unsupervised Learning because there is no supervision  $y$

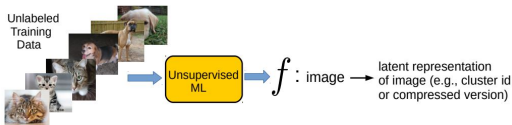


# Function Approximation in Machine Learning

- **Supervised Learning** (“predict  $y$  given  $x$ ”) can be thought learning a function that maps  $x$  to  $y$

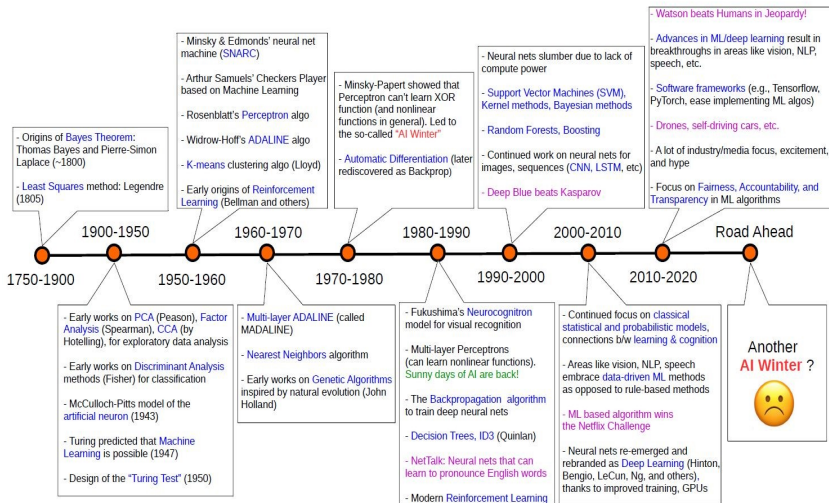


- **Unsupervised Learning** (“model  $x$ ”) can also be thought of as learning a function that maps  $x$  to some useful latent representation of  $x$



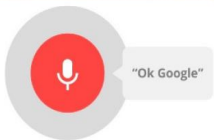
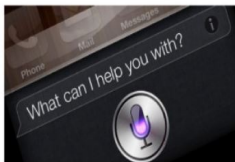
- Other ML paradigms (e.g., Reinforcement Learning) can be thought of as doing function approximation.

# Machine Learning: A Brief Timeline and Some Milestones



# Machine Learning in the real-world

Broadly applicable in many domains (e.g., internet, robotics, healthcare and biology, computer vision, NLP, databases, computer systems, finance, etc.).



Predictive Policing



Online Fraud Detection

# Machine Learning helps Natural Language Processing

ML algorithm can learn to translate text



(even "transliterate")

# Machine Learning meets Speech Processing

## ML algorithms can learn to translate speech in real time

**PUTTING MACHINE LEARNING TO THE TEST**  
To provide a seamless user experience, Skype Translator uses machine learning to solve key challenges in interpreting human language, including:

-  **Representing the different ways people really speak**
-  **Determining sentence boundaries, punctuation and ease from speech**
- there they're their**  
 **Disambiguating sound-alike words in context**
-  **Mapping words and phrases from one language to another**



### NOW YOU'RE SPEAKING MY LANGUAGE (LITERALLY)



Skype has always been about making it easy to talk with family and friends all over the world. Now, by integrating advanced speech recognition and automatic translation into Skype, Skype Translator lets you speak with those you've always wished you could, even if they speak a different language.

#### HOW SKYPE TRANSLATOR WORKS

Automatic Speech Recognition	Speech Correction	Translation	Text to Speech	Using and Teaching the system
 <p>"Hi, Grandma! I am so excited to speak to you!"</p>	 <p>"s-so excited, ah..." "so excited, ah..." "so excited..."</p> <p>A deep neural network analyzes spoken words against audio snippets from millions of previously recorded samples and transforms the audio to a set of text candidates.</p>	 <p>English ↔ Español</p> <p>Skype Translator has learned how dozens of languages align with one another by reviewing millions of pairs of previously translated content. Using Microsoft Translate, the same tool used in numerous Microsoft products, it applies this knowledge to quickly translate the text into Spanish.</p>	 <p>"Hola, abuelita! Estoy muy emocionada de hablar con usted!"</p>	 <p>using the system teaching the system</p> <p>Increased usage and user feedback, plus constant refinement by human translators, help Skype Translator learn and get better.</p>

#### TRANSLATE INSTANT MESSAGES IN OVER 40 LANGUAGES

Having a translated IM conversation is super easy! Choose a contact, turn on the Translation switch for that person, and start typing. When you're "too slow", your original message will appear in the right-hand pane, followed by its translation. Your contact on the other end will see something very similar, albeit with the translated message in his/her preferred language presented first. While voice translation initially supports English and Spanish only, full translation supports over 40 languages, so feel free to experiment with them all...even Klingon!



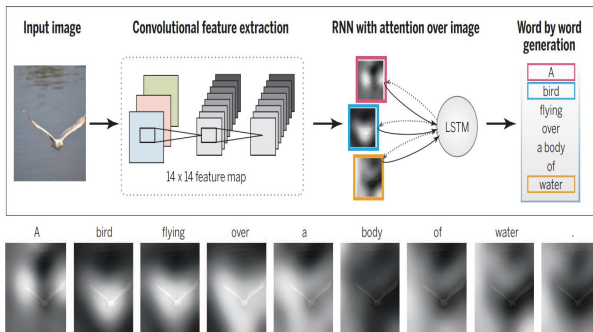
Register for the preview at [www.skype.com/translator](http://www.skype.com/translator) and wait for your invite.  
Install the Skype Translator client.  
Use Skype Translator to call someone who speaks Spanish. Or, if you speak Spanish, call someone who speaks English.  
Every call you make helps Skype Translator get a little bit better. You won't see the improvement right away, but you will see gradual improvement over time.

# Machine Learning helps Computer Vision

- Automatic generation of text captions for images:

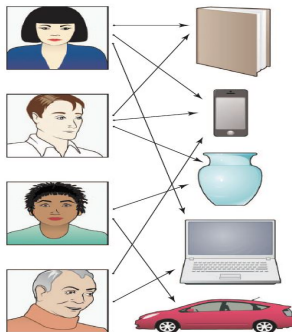
A **convolutional neural network** is trained to interpret images, and its output is then used by a recurrent neural network trained to generate a text caption.

- The sequence at the bottom shows the word-by-word focus of the network on different parts of input image while it generates the caption word-by-word.



# Machine Learning helps Recommendation systems

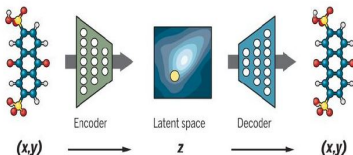
- A **recommendation system** is a machine-learning system that is based on data that indicate links between a set of users (e.g., people) and a set of items (e.g., products).
- A link between a user and a product means that the user has indicated an interest in the product in some fashion (perhaps by purchasing that item in the past).
- The **machine-learning problem** is to suggest other items to a given user that he or she may also be interested in, based on the data across all users.



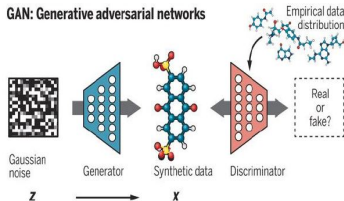
# Machine Learning helps Chemistry

ML algorithms can understand properties of molecules and learn to synthesize new molecules<sup>1</sup>.

## VAE: Variational autoencoders

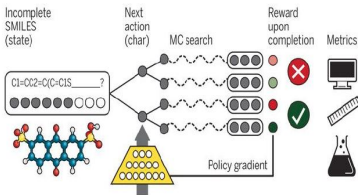


## GAN: Generative adversarial networks

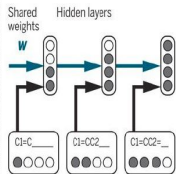


## RL: Reinforcement learning

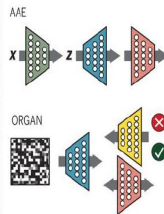
Policy gradient with Monte Carlo tree search (MCTS)



## RNN: Recurrent neural network



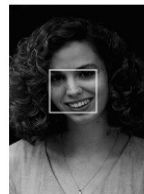
## Hybrid approaches



<sup>1</sup>Inverse molecular design using machine learning: Generative models for matter engineering (Science, 2018)

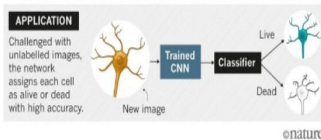
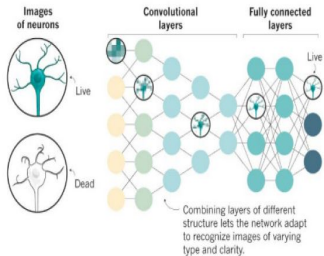


# Machine Learning helps Image Recognition



# Machine Learning helps Many Other Areas...

## Biology



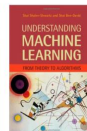
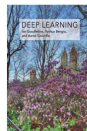
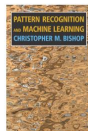
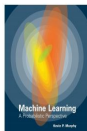
## Finance



Start with a book of 150 Pages



Then you can start reading these books



- <http://www.learningtheory.org/>
- <https://www.kdnuggets.com/>
- <http://archive.ics.uci.edu/ml/index.php>
- <https://sebastianraschka.com/resources.html>
- <http://cs229.stanford.edu/syllabus-spring2019.html>
- <https://www.ctanujit.org/lecture-notes.html>

Questions?

# Thank You