Overview

Suleyman Demirel University

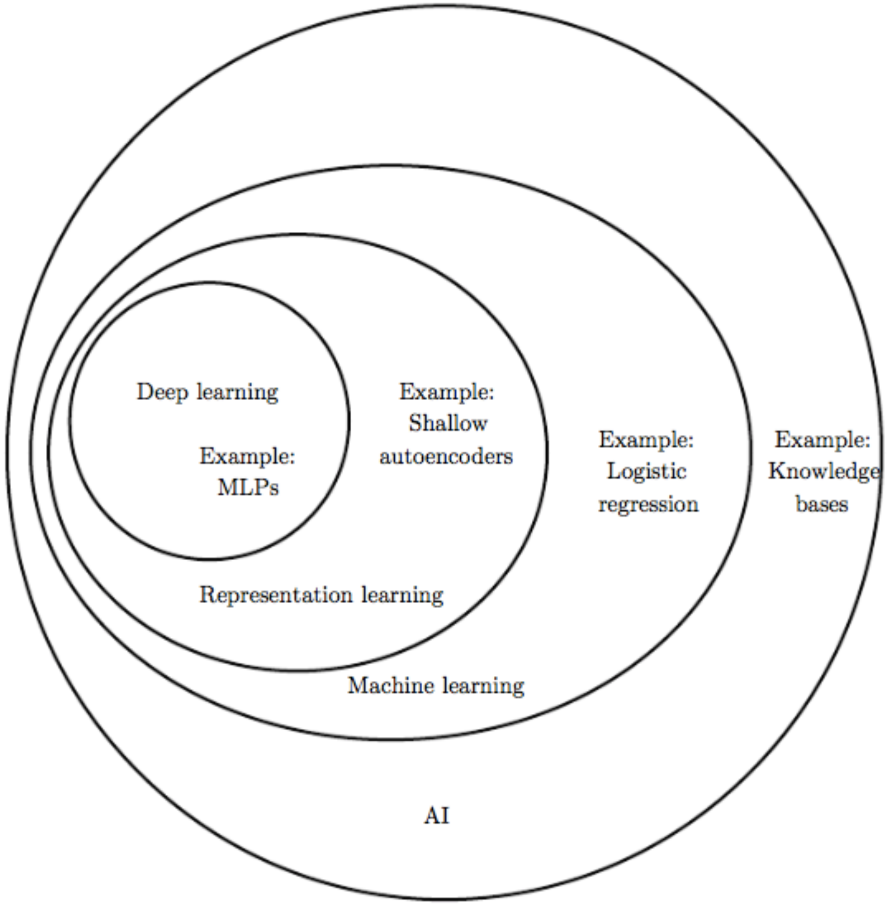
CSS634: Deep Learning

PhD Abay Nussipbekov

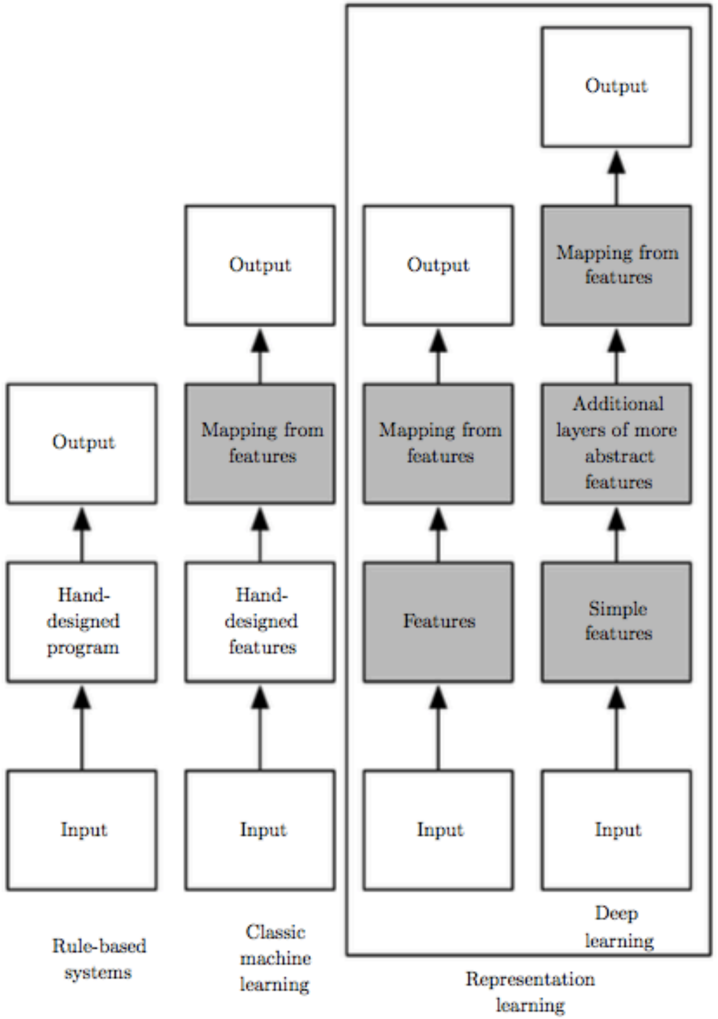
Topics

- Overview
- The Perceptron
- Logistic Regression and Multi-class Classification
- Neural Networks
- Regularization, Normalization and Weight Initialization
- Learning Rates and Optimization Algorithms
- Convolutional Neural Networks
- Recurrent Neural Networks
- Autoencoders
- Generative Adversarial Networks
- Reinforcement Learning

What is Deep Learning?

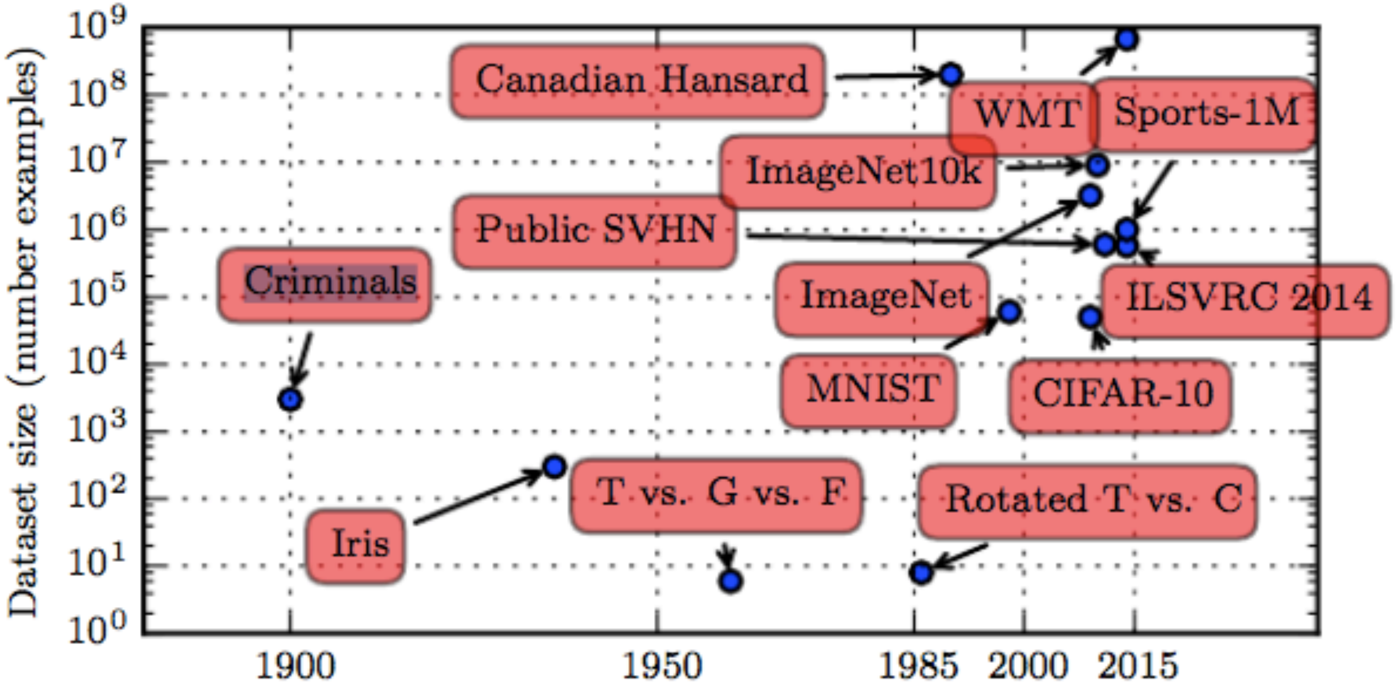


What is Deep Learning?



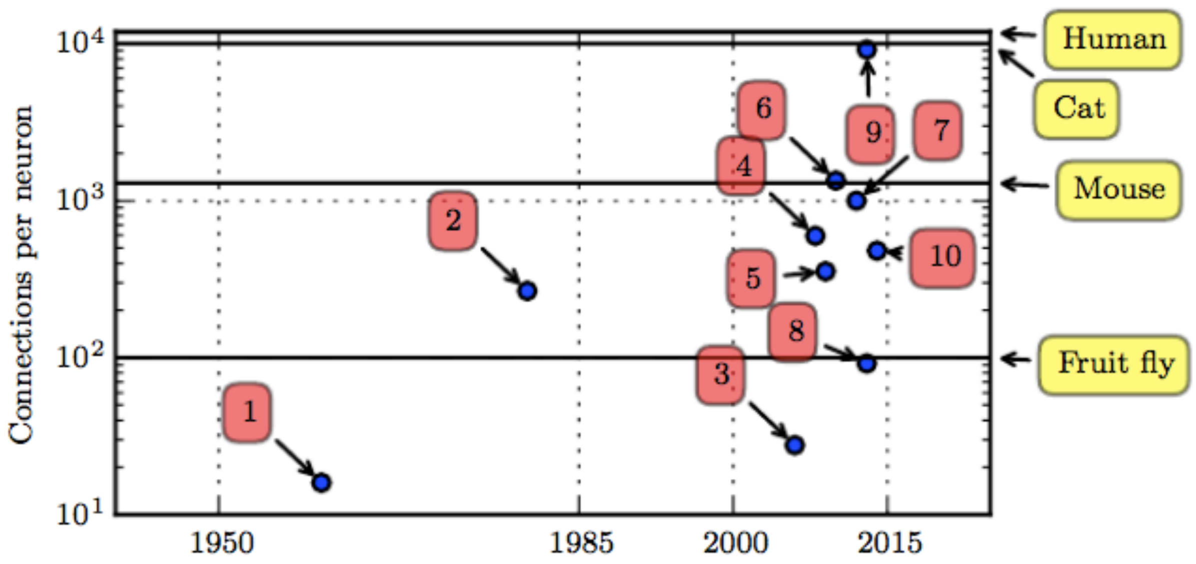
Increasing Dataset Sizes

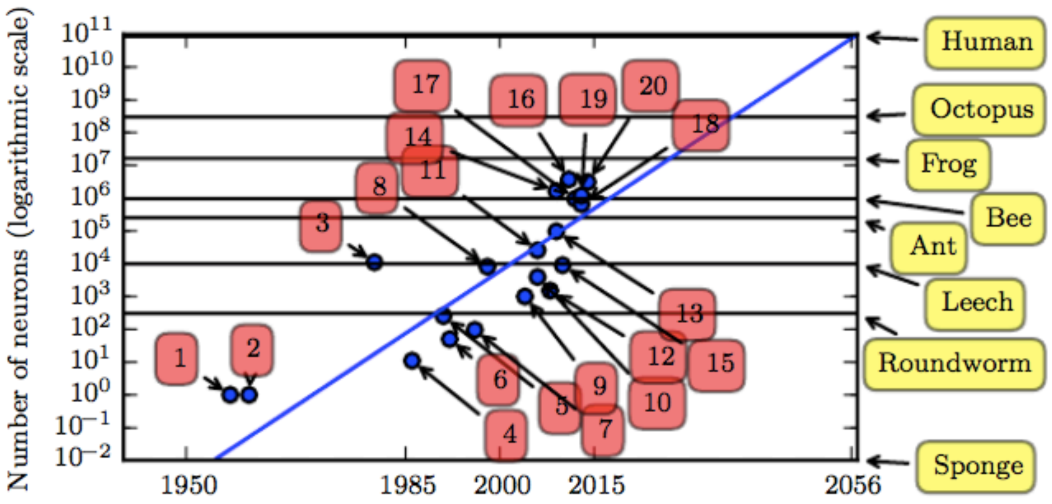
- Acceptable performance with around 5,000 labeled examples per category
- Human performance when at least 10 million examples



Increasing Model Sizes

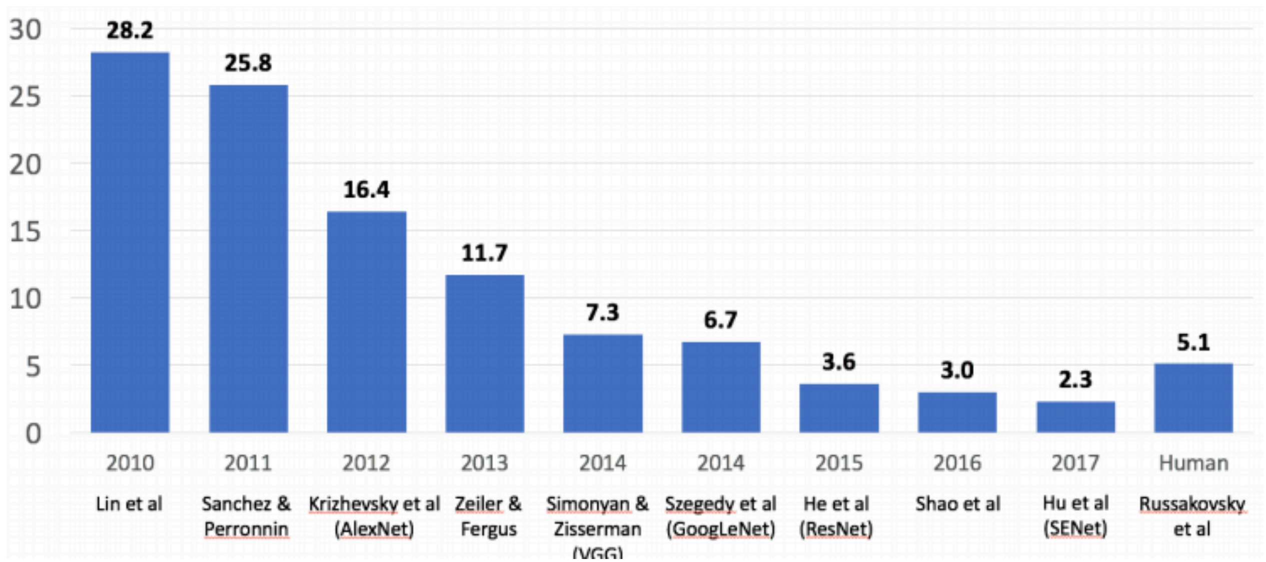
- Unless new technologies allow faster scaling, we will not have the same number of neurons as the human brain until at least the 2050s (may be even larger as biological neurons may represent more complicated functions).

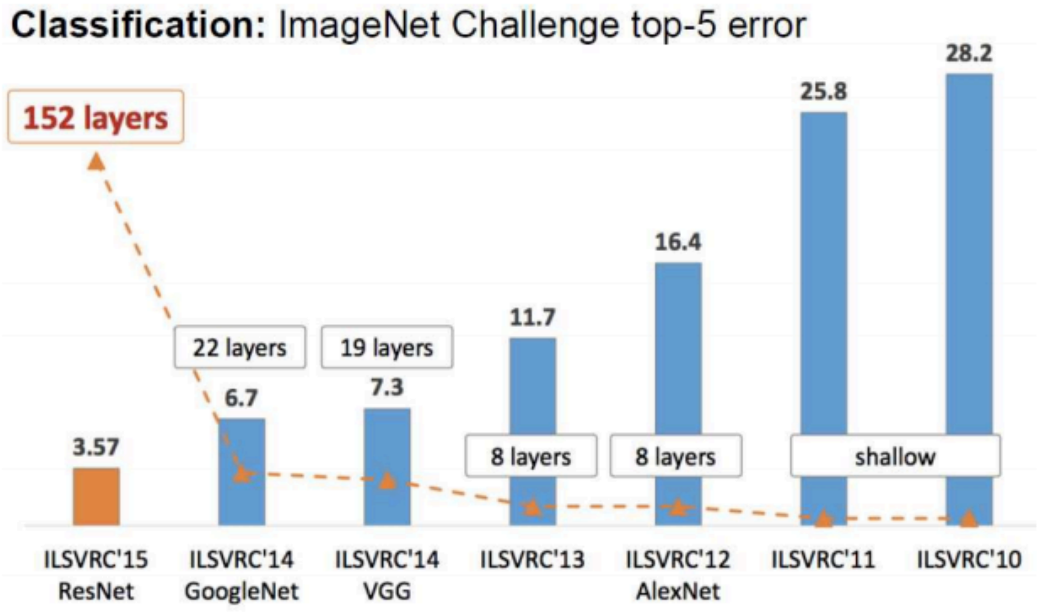




Increasing Accuracy, Complexity and Real-World Impact

- Individual objects in tightly cropped, extremely small images (Rumelhart et al., 1986a)
- Modern networks typically recognize at least 1,000 different categories of objects




Classification: ImageNet Challenge top-5 error

AI Effect

AIS researcher Rodney Brooks complains "Every time we figure out a piece of it, it stops being magical; we say, 'Oh, that's just a computation.'"

Before diving in

Let’s recap what is machine learning first…

Machine Learning

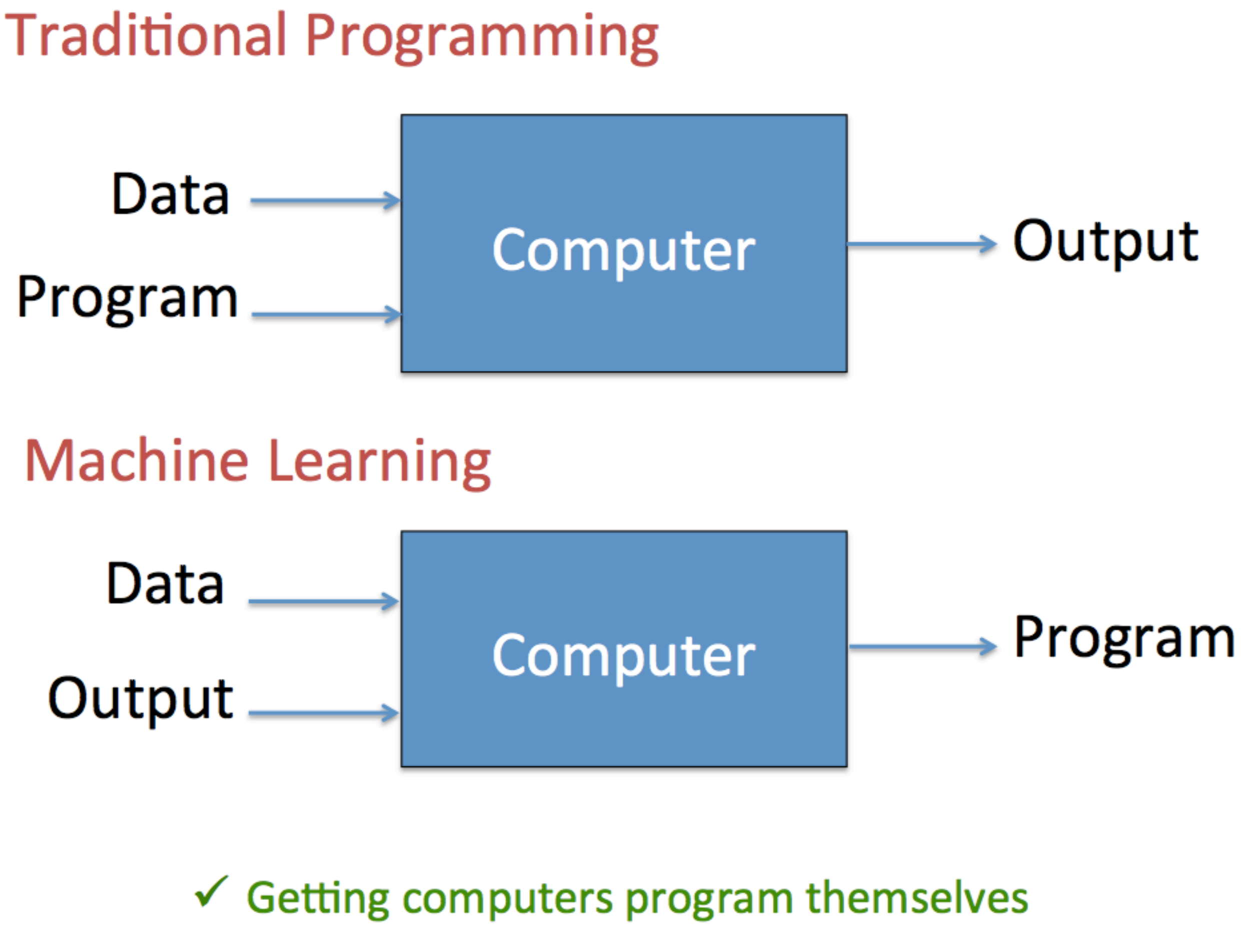
- “Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed” – Arthur L. Samuel, AI pioneer, 1959

- “Machine learning is automating automation” (Prof. Pedro Domingo, University of Washington)

Why Machine Learning?

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, former director, DARPA)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, former CTO, Sun)
- Machine learning today is one of the hottest aspects of computer science” (Steve Ballmer, CEO, Microsoft)

Machine Learning



Traditional Programming

Data

Program

Computer

Output

Machine Learning

Data

Output

Computer

Program

✓ Getting computers program themselves

Applications of ML/DL

- Email spam detection
- Face identification, recognition, verification (e.g., iPhone X)
- Web search (e.g., Google, Baidoo)
- Sports analytics and predictions
- ATMs (e.g., reading checks)
- Credit card fraud
- Stock predictions

Applications of ML/DL

- Smart assistants (Apple Siri, Amazon Alexa, ...)
- Product recommendations (e.g., Netflix, Amazon)
- Self-driving cars (e.g., Uber, Tesla)
- Language translation (Google translate)
- Robotics
- Drug design
- Medical diagnoses
- Space
- …

Types of Learning

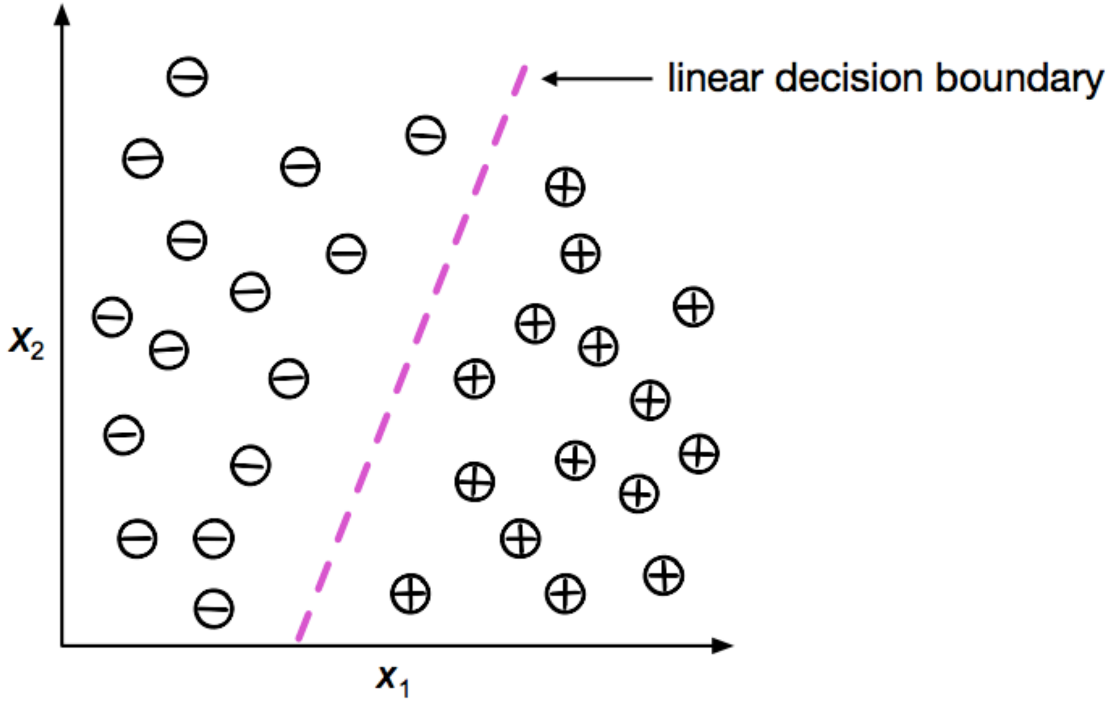
- Supervised learning
  - Training data includes desired outputs
  - Classification, regression
- Unsupervised learning
  - Training data does not include desired outputs
  - Find hidden structure in data
  - Clustering, dimensionality reduction/representation learning
- Reinforcement learning
  - The learner interacts with the world via “actions” and tries to find an optimal policy of behavior with respect to “rewards”(feedback) it receives from the environment
  - Learning with a critic who merely states that something is right or wrong, but does not say specifically how it is wrong.

Some Other Types of Learning

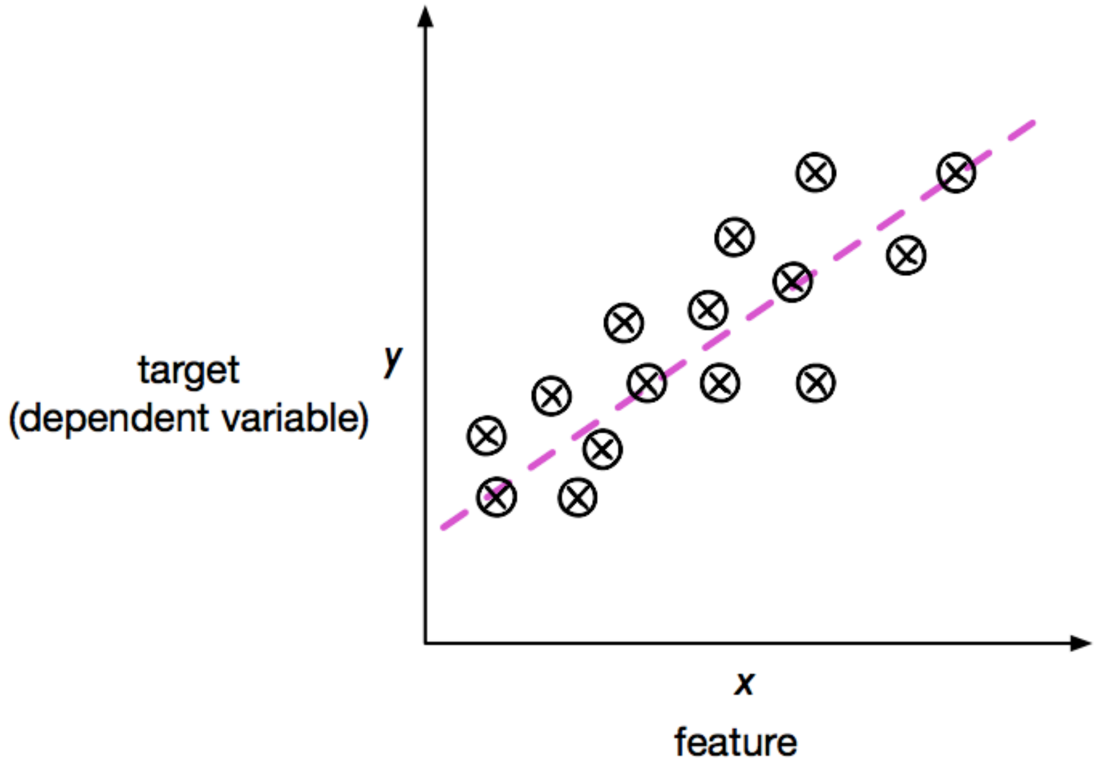
- Semi-supervised learning
  - Training data includes a few desired outputs
- Self supervised learning
  - Type of supervised learning where training data is autonomously labeled
  - In self-supervised learning, the system learns to predict part of its input from other parts of it input. In other words a portion of the input is used as a supervisory signal to a predictor fed with the remaining portion of the input. (Yann LeCun)

Supervised Learning: Classification

Binary classification example with two features ("independent" variables, predictors)

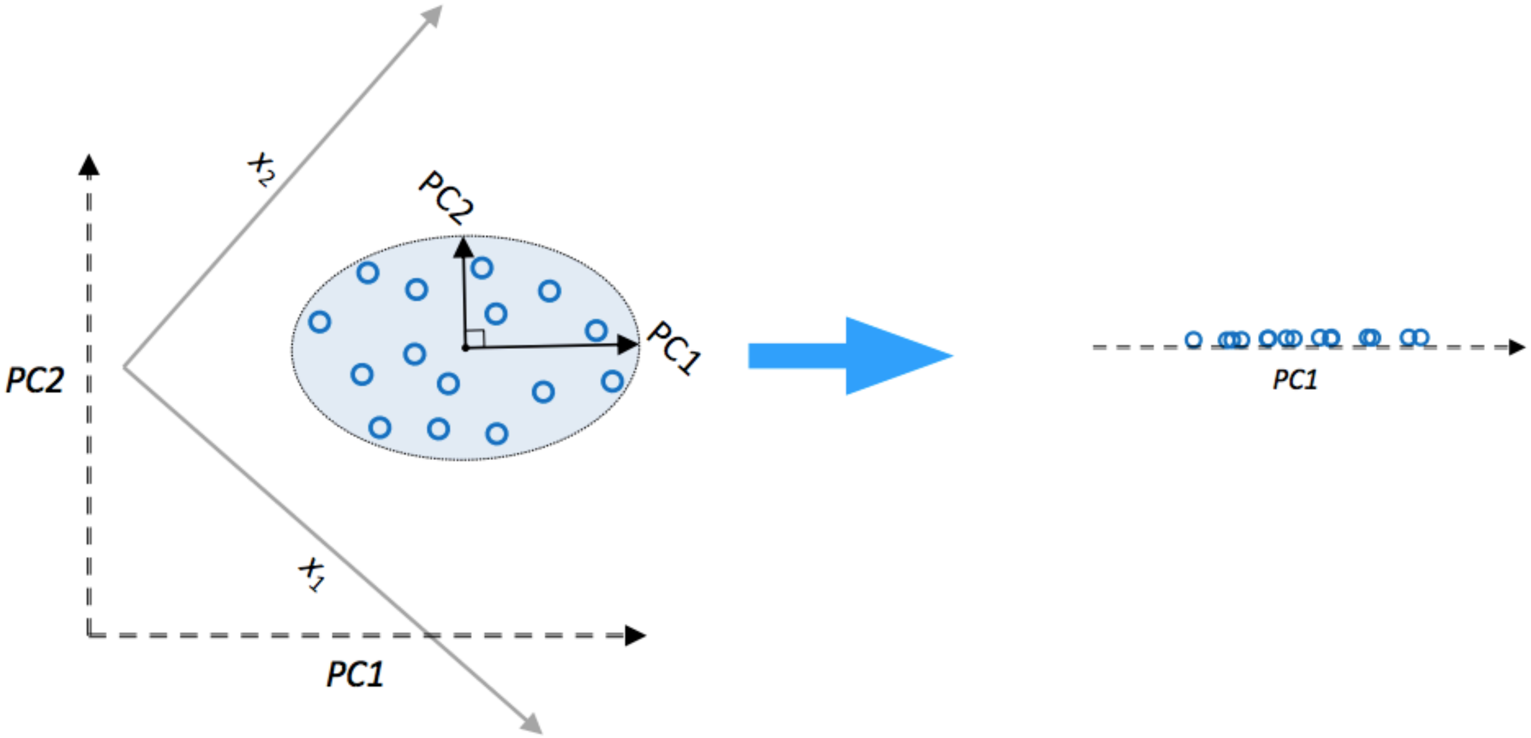


Supervised Learning: Regression



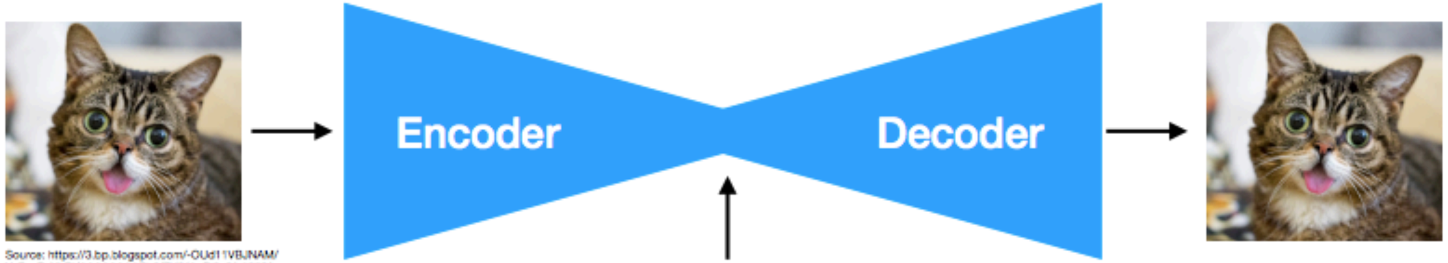
Unsupervised Learning: Representation Learning/ Dimensionality Reduction

E.g. Principal Component Analysis



Unsupervised Learning: Representation Learning/ Dimensionality Reduction

E.g. Autoencoders



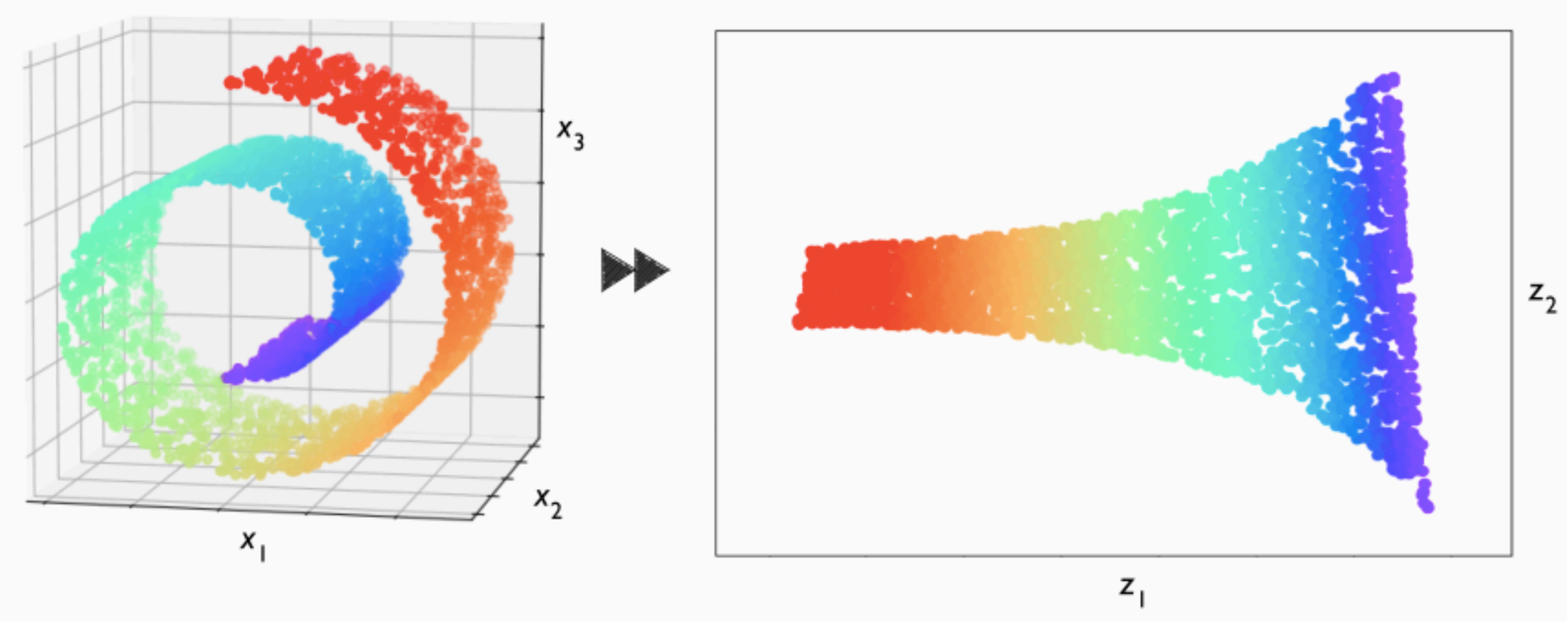
Source: https://3.bp.blogspot.com/-OUd11VBJNAM/VsFacR_YhBI/AAAAAAAABh0/ZKfKAnRj3x0/s1600/cannot%2Bresist.jpg

Encoder

Decoder

latent representation/
feature embedding

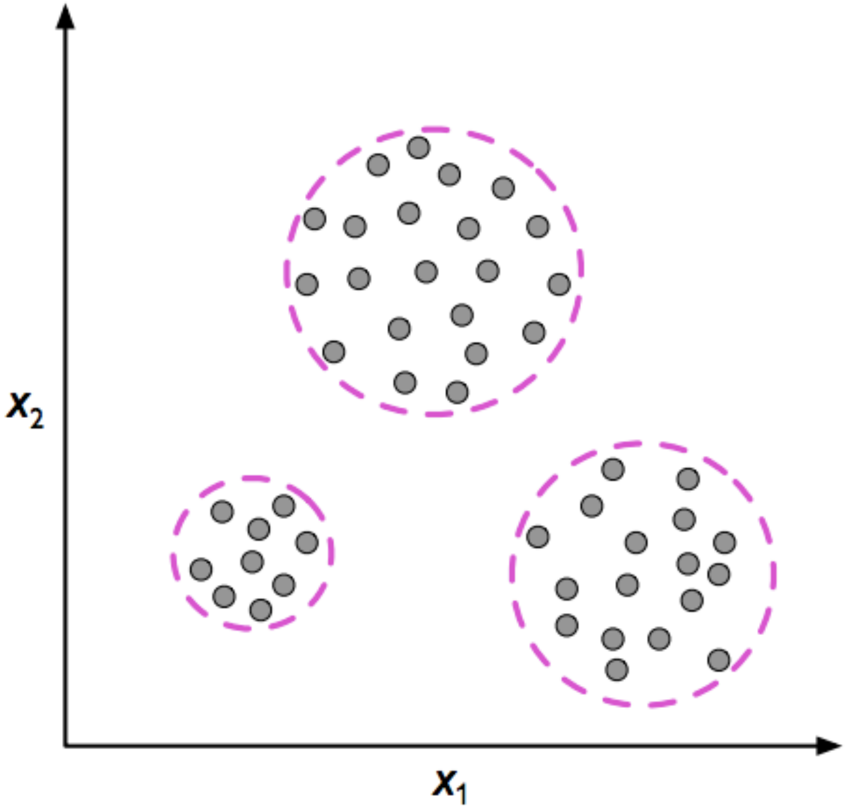
Unsupervised Learning: Representation Learning/ Dimensionality Reduction



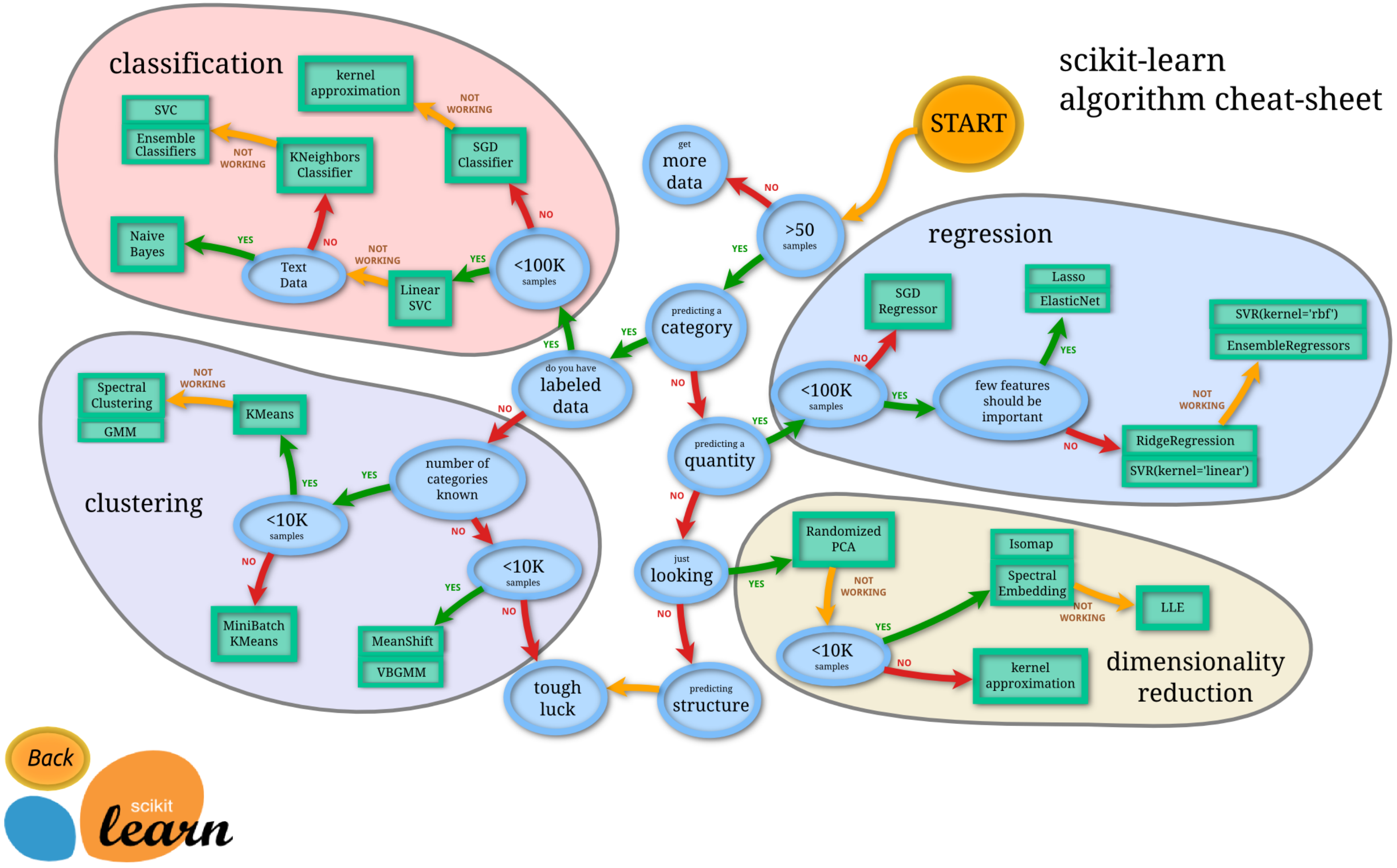
Example of manifold learning using kernel PCA

Unsupervised Learning: Clustering

Assigning group memberships to unlabelled examples

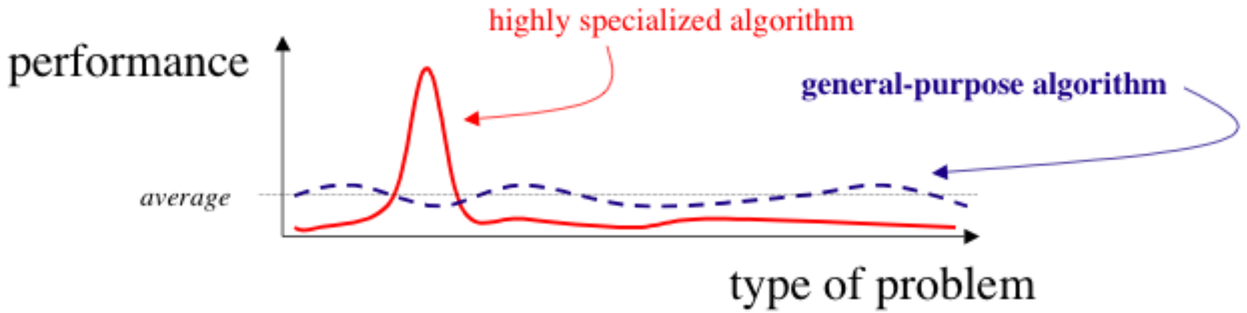


How to Choose Right Algorithm?



The No Free Lunch Theorem

- "Any two optimization algorithms are equivalent when their performance is averaged across all possible problems" (Wolpert 1996).
- In other words, in some sense, no machine learning algorithm is universally any better than any other.



Steps in developing ML application

1. *Collect data*
   - Scraping a website
   - RSS feed
   - API
   - Read from different devices (wind speed, blood glucose levels, etc.)
   - Use publicly available data
2. *Prepare the input data*
   - Make sure it’s in a usable format (lists, matrices, dictionaries, etc.)
   - Algorithm-specific formatting (e.g. need features in a special format, target variables and features as strings or integers)
   - The choice of the distinguishing features is critical. Depends on problem. Needs prior knowledge.

Steps in developing ML application

3. *Analyze the input data*
   - Looking at the data simply in a text editor to make sure you don't have empty values or if you can recognize any patterns or if there's anything obvious, such as a few data points that are vastly different from the rest of the set.
   - Plotting data in one, two, or three dimensions can help. But most of the time you'll have more than three features.
   - If it's a production system and you know what the data should look like, or you trust its source, you can skip this step. This step is human involvement which is not for automated system. It makes you understand you don't have garbage coming in.

Steps in developing ML application

4. *Train the algorithm*
   - This is where the ML takes place. This step and the next step are where the “core” algorithms lie, depending on the algorithm. You feed the algorithm good clean data from the first two steps and extract knowledge or information.
   - Unsupervised algorithms skip this step

Steps in developing ML application

5. Test the algorithm
   - Evaluate it to see how well it does. In the case of supervised learning, you have some known values you can use to evaluate the algorithm. In unsupervised learning, you may have to use some other metrics to evaluate the success.
   - In either case, if you’re not satisfied go to step 4 or 1
   - Cost – depends on problem domain.

Main components of ML

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
  - Representation
  - Evaluation
  - Optimization

Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

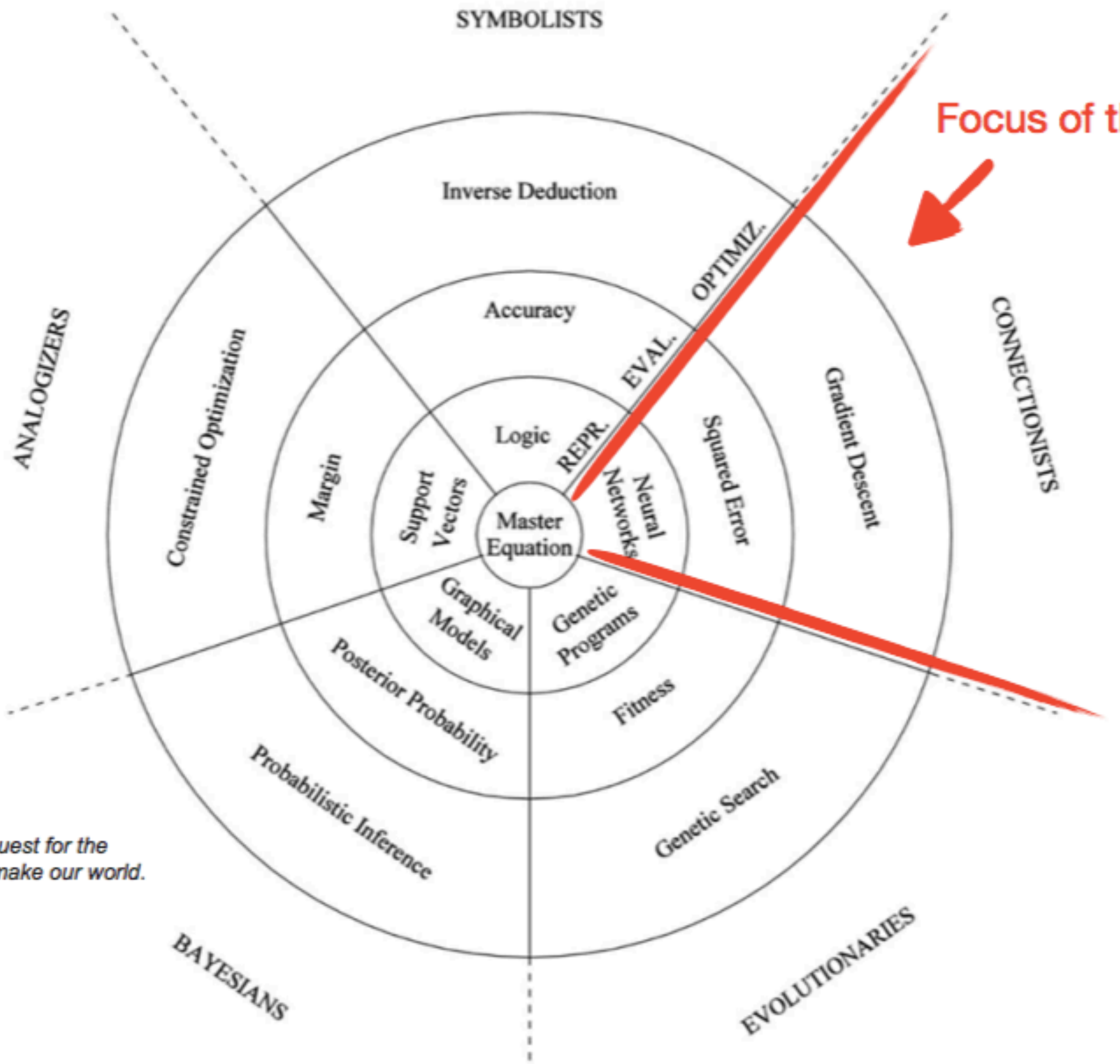
Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost/Utility
- Margin
- Entropy
- Etc.

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

Pedro Domingo's 5 Tribes of Machine Learning



Source: Domingos, Pedro.
The Master Algorithm: How the quest for the ultimate learning machine will remake our world.
Basic Books, 2015.

Cost Functions

- Maximize the posterior probabilities (e.g., naive Bayes)
- Maximize a fitness function (genetic programming)
- Maximize the total reward/value function (reinforcement learning)
- Maximize information gain/minimize child node impurities (CART decision tree classification)
- Minimize a mean squared error cost (or loss) function (CART, decision tree regression, linear regression, adaptive linear neurons, ...)
- Maximize log-likelihood or minimize cross-entropy loss (or cost) function
- Minimize hinge loss (support vector machine)

Resources Used

- Relatively new field, no “best” textbook
- Deeplearningbook by Ian Goodfellow, Yoshua Bengio and Aaron Courville
- Deeplearning.ai course by Andrew Ng
- CS231n: Convolutional Neural Networks for Visual Recognition by Fei-Fei Li, Andrej Karpathy, Justing Johnson
- STAT 479: Deep Learning by Sebastian Raschka
- CMSC 35246 Deep Learning by Shubhendu Trivedi and Risi Kondor
- Scientific papers, online references and etc.

Grade Policy

Midterm exam:	30%
Assignments:	30%
Final	
- Practical part:	20%
- Theoretical part:	20%

Recommended Background

Linear Algebra: vectors, matrices, linear equations, ...

Sanity check: What is identity matrix? Lp norm?

Probability and Statistics: types of probabilities, distributions, Bayes theorem, likelihood

Sanity check: Probability vs likelihood? Covariance matrix? pdf? Bayes theorem? Why is "naïve Bayes" naive?

Information Theory: entropy, ...

Sanity check: why KL divergence is not a distance measure between 2 vectors?

Calculus: derivatives, chain rule, ...

Sanity check: Chain rule? Jacobian matrix?

Numerical Calculations: gradient descent, ...

Sanity check: what is numerical underflow?

Machine Learning: representation, evaluation, optimization

Sanity check: How to choose appropriate algorithm? How to detect overfitting and underfitting? Evaluation? Bootstraping, bagging, boosting?

Algorithms and Data Structures: time and memory complexities, dictionaries, sets, lists, sorting, ...

Sanity check: explain hash map

Programming proficiency in Python:

Sanity check: Given two sparse matrices A and B, return the result of AB. (Assume A's column number is equal to B's rows number)

```
[p for p in range(100) if not (x for x in range(2,p) if not p%x)]
```

What are the generators and iterator in Python?

Confidence in Numpy:

Sanity check: how to find the positions of missing values in numpy array?

Desire: if you don't have any of the above but a really strong desire, consistency and patience then that's enough!