Machine Learning I

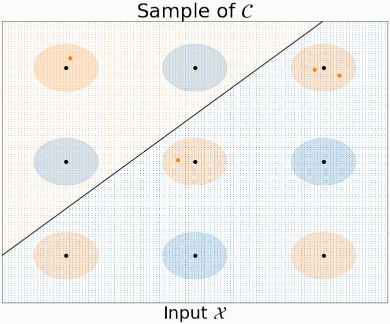
Lecture 22: Survey Of Further Directions

Nathaniel Bade

Northeastern University Department of Mathematics

Table of contents

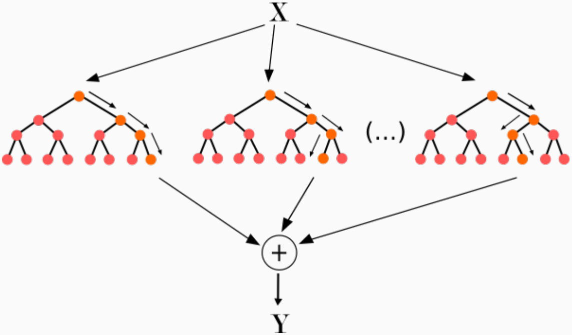
What We Have Done



We started this semester by answering the question of what it means, mathematically, for an unknown distribution to be learnable by a class of function. We derive strong negative results like the No Free Lunch theorem, explanatory relations like the Bias-Complexity trade-off, and learnability criteria, like the VC dimension and the PAC learning framework.

Characteristic	Neural Nets	SVM	Trees	MARS	k-NN, Kernels
Natural handling of data of “mixed” type	▼	▼	▲	▲	▼
Handling of missing values	▼	▼	▲	▲	▲
Robustness to outliers in input space	▼	▼	▲	▼	▲
Insensitive to monotone transformations of inputs	▼	▼	▲	▼	▼
Computational scalability (large N)	▼	▼	▲	▲	▼
Ability to deal with irrelevant inputs	▼	▼	▲	▲	▼
Ability to extract linear combinations of features	▲	▲	▼	▼	◆
Interpretability	▼	▼	◆	▲	▼
Predictive power	▲	▲	▼	◆	▲

In the next portion of the class, we moved discussed algorithms, did exact statistical calculations for bias and variance for linear and quasi-linear methods, and finally canvased the space of modern algorithm in machine learning. In addition to supervised learning, we talked about clustering and dimensional reduction, and how information preservation leads to different kinds of encodings.



Finally, in the last week or so, we have discussed moving beyond single functions and straight forward learning to incorporate black box methods: Smoothing and basis expansion, bootstrapping and bagging, and finally hyperparamter tuning. This brings us to the forefront of machine learning, what could possibly be left?

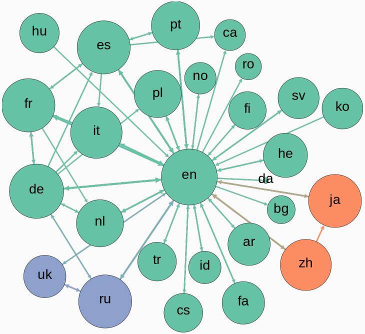
Topics We Missed

Topics We Missed

Machine learning is a large, cross disciplinary field and although we've given a survey of many of the main aspects we have left out enough for at least another full course. Each field that uses machine learning, from social science, to biology, chemistry and physics, to mechanical and industrial engineering has it's own set of tools tactics, and techniques.

Many of these tools build on the algorithms and structures we discussed in this course, but some start from a completely different framework, or cut up the problem space in unique ways.

Graphs and Networks



Graphs are one of the basic data structures in mathematics. They are also a intuitive tool for understanding relationships. From social networks to state machines, network analysis has become one of the most important developing fields in the 21'st century.
(Network graph formed by Wikipedia editors (edges) contributing to different Wikipedia language versions (vertices) during one month in summer 2013.)

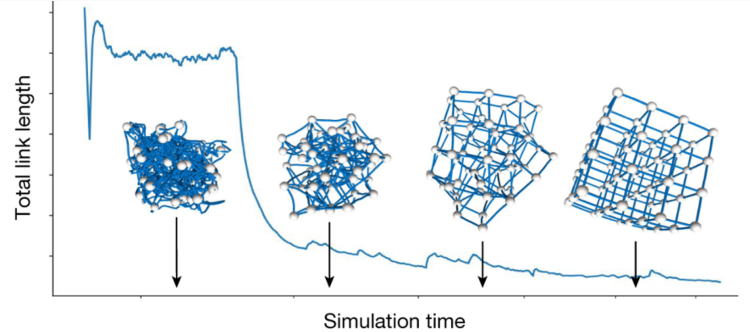
Graphs and Networks

Network analysis is used to understand political affiliation, social relationships, as well as modeling physical phonominia like molecular bonding. Machine learning on graphs can be used to grow and prune relational dataset, predict graph structures or learn nontrivial local topology.

Graphs and Networks

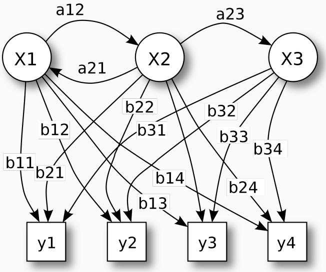
For example, in “Fake news on Twitter during the 2016 U.S. presidential election”, grouping was performed using an ensemble of clustering algorithms to decide which sites represent legitimate news sources, and which represent serial publishers of fake news.

Graphs and Networks



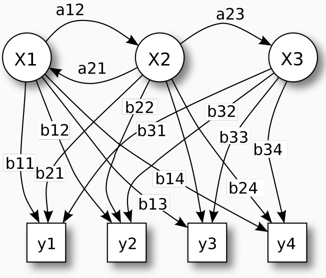
Machine learning is also used to construct graphs models. In the physical models above, physical constants for molecular structures are fit by generating graphical models and evaluating their energy based on linkage length.

Hidden Markvo Models



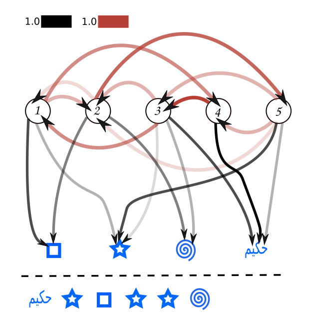
Hidden Markov models are mathematical state machines on which only the output can be seen. Unlike Markov chains, where at each step the true state is available to the viewer, in a HMM we assume only the output is known, and try to infer the sequence off internal states from the result.

Hidden Markvo Models



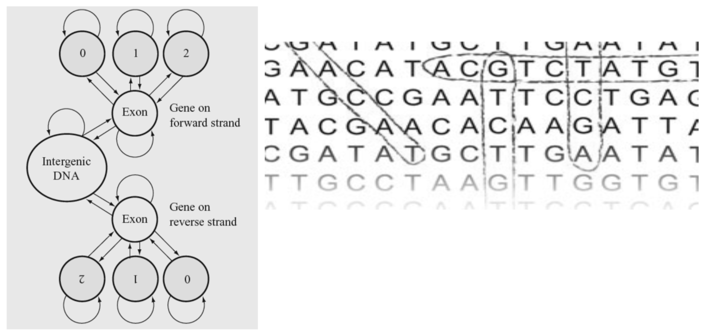
In the above, x_i are states, y_j are results, $a_{ii'}$ are state to state transition probabilities and *bij* are the probabilities of a model in state x_i outputting state y_j. In modeling something as a HMM, we assume an internal number of nodes and try to learn the probabilities $a_{ii'}$ and b_{ij} from the output.

Hidden Markvo Models



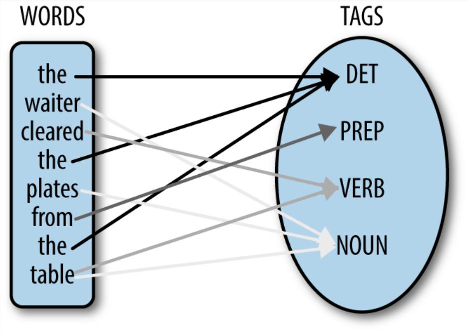
Hidden Markov models model black box phonomina. They have been used as models in cryptanalysis, speech recognition (including Siri),

Hidden Markvo Models



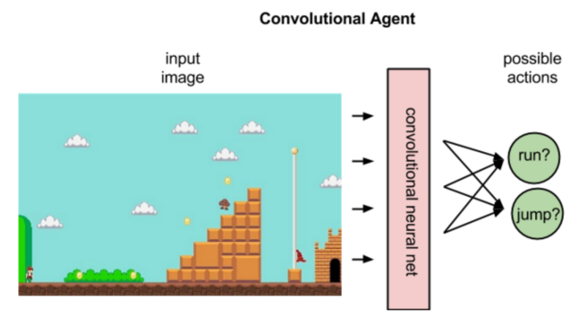
Hidden Markov models model black box phonomina. They have been used as models in cryptanalysis, speech recognition (including Siri), genetic analysis,

Hidden Markvo Models



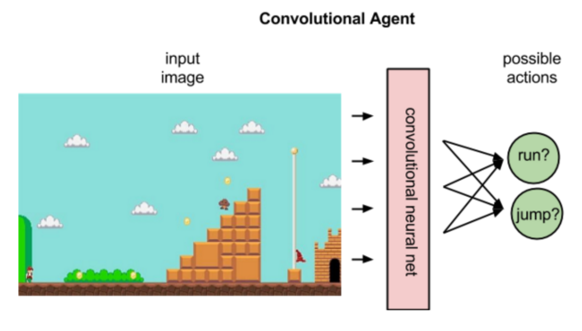
Hidden Markov models model black box phonomina. They have been used as models in cryptanalysis, speech recognition (including Siri), genetic analysis, and NLP.

Reinforcement Learning

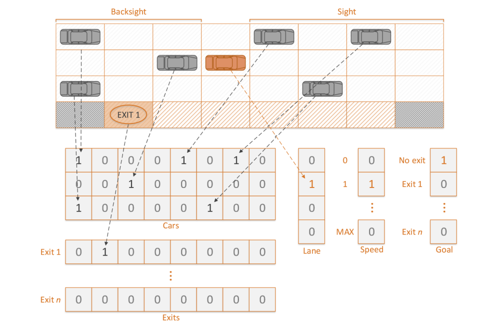


Reinforcement learning refers to a paradigm of machine learning and agent based modeling where a software agent takes actions on an **interface** based on some **environment** to maximize some notion of **reward**. Reinforcement learning is probably what is most traditionally thought of as AI in the modern context.

Reinforcement Learning

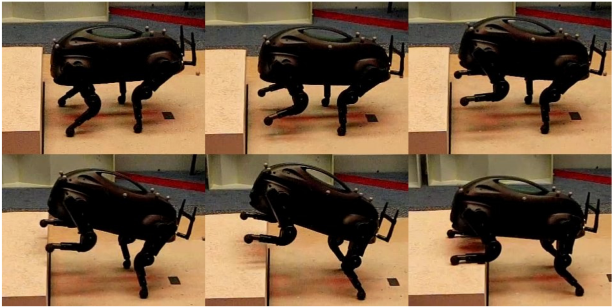


Mathematically, reinforcement learning trades a classifier for a **policy**, which tells the agent how to update at each step. There are many reinforcement learning algorithms, often in tight relation to tactics in classification learning problems.

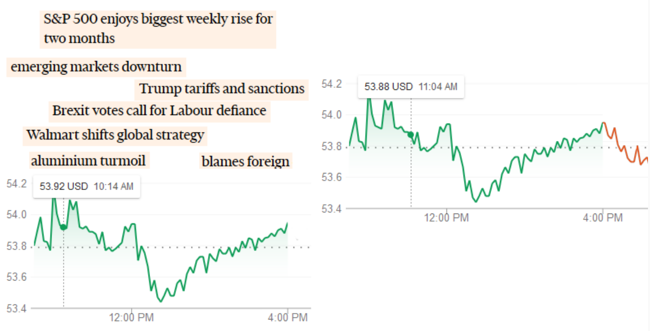


Reinforcement learning traditionally differs from static classification in that it is a type of **online learning**. Online learning problems actively learn as new information is presented. Online reinforcement learning has found application in stock trading, self driving cars, and partial control mechanisms for computer guided robotics. (Image Source)

Reinforcement Learning

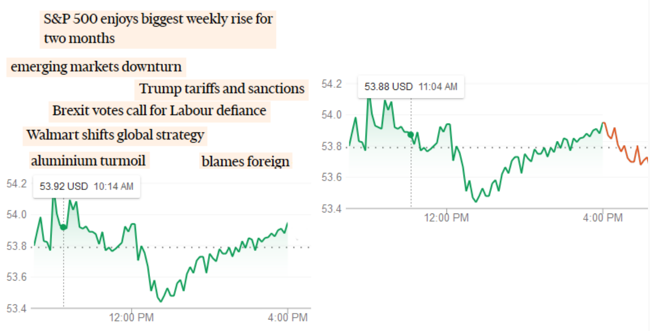


Reinforcement learning traditionally differs from static classification in that it is a type of **online learning**. Online learning problems actively learn as new information is presented. Online reinforcement learning has found application in stock trading, self driving cars, partial control mechanisms for computer guided robotics. (Image Source)



Finally, the next step beyond the single model paradigms we've discussed in this course is into the space of **ensemble learning**. We've discussed ensambles of a single type of model on a fixed dataset, but part of the power of machine learning is the ability to simultaneously fit different types of models to different types of datasets, all leading towards a single reward.

Ensemble Learning



The efficient simultaneous training and processing of multiple types of data, using multiple algorithms, is one of the biggest challenges in modern machine learning. Reaching across industries and academic disciplines.

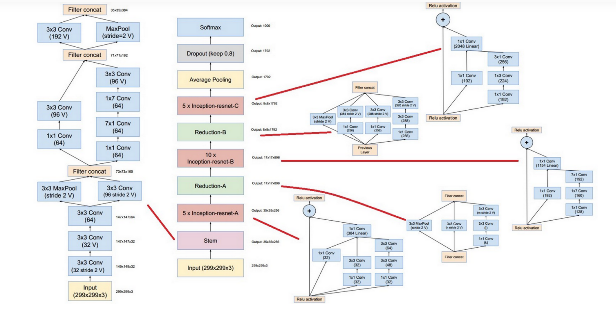
Ensemble Learning

In 2009, the Netflix movie recommendation prize was won by "BellKor's Pragmatic Chaos", a merger of teams "Bellkor in BigChaos" and "Pragmatic Theory." Their only close competitors were another merger of teams, "The Ensemble," both achieve a 10.09% improvement over Netflix own algorithm. Of all of the constituent teams, the best score had been an 8.43% improvement.

Ensemble Learning

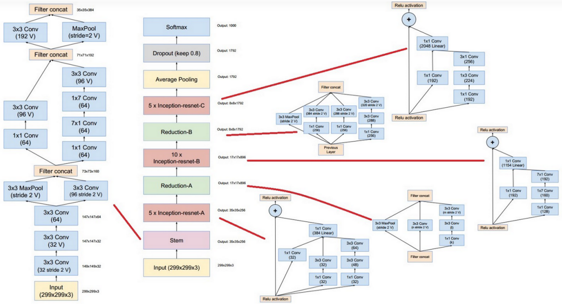
The Netflix prize ushered in a new paradigm in production level machine learning: single models can be good, but ensambled models are king. (Image Source)

Ensemble Learning



Inception-ResNet-V2 Architecture

Finally, within ensembling and hyperparameter tuning lies the future of large scale machine learning: **transfer learning** is the paradigm of creating pretrained models that can be snapped together for last mile ensamble training.



Inception-ResNet-V2 Architecture

There is a race to develop these kinds of snap together architecture, and examples like Google's Inception net above are already in use for picture classification. In natural language processing, the groundbreaking (ULM-FiT) paper of 2018 showed that pretrained models could learn new tasks with as a few as 100 examples, out performing by up to 24% models trained on datasets in the 10,000's.

Where To Go From Here

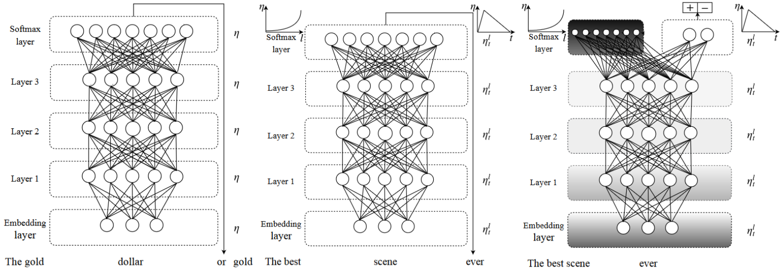
Big Ideas and Further Directions

In this course we've talked mostly (and a bit too much) about models as abstracted from problems. I want to take a moment and discuss some of the big developments in machine learning in the last year, with applications to:

- Natural Language Processing.
- Computer Vision.
- Instability Analysis.
- Visualization.

In general, embedding and low dimensional representation has dominated the most celebrated recent results. A low dimensional embedding not only greatly speeds training, but has a hope of holding rich information about the structure of data.

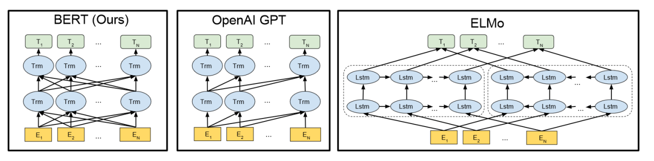
Natural Language Processing



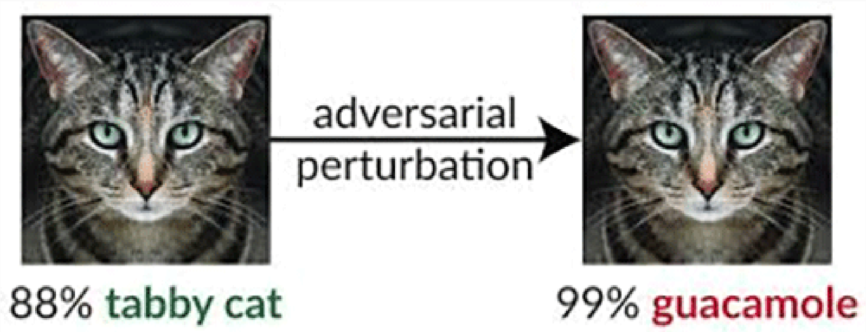
(a) LM pre-training (b) LM fine-tuning (c) Classifier fine-tuning

Natural language processing is one of the biggest topics in machine learning, and many of the advances in the subject have come from world of translation, textual processing and prediction from language. ULMFiT featured showed that transfer learning could speed up text processing on a variety of tasks. ULMFiT has been applied to multiple languages, including German, Polish, Hindi, Indonesian, Chinese, and Malay with drastic improvements.

Natural Language Processing



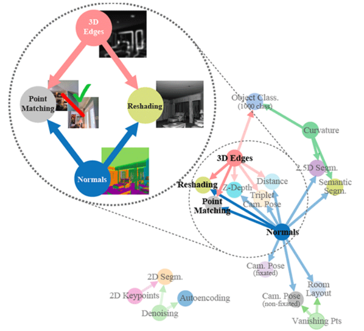
In addition, unsupervised learning provides reduction and storage mechanisms for language that many people think are the key to universal machine translation. Both the Allen Institutes ELMo and Google AI's BERT use intelligent language encoding to drastically speed up NLP tasks and improve accuracy.



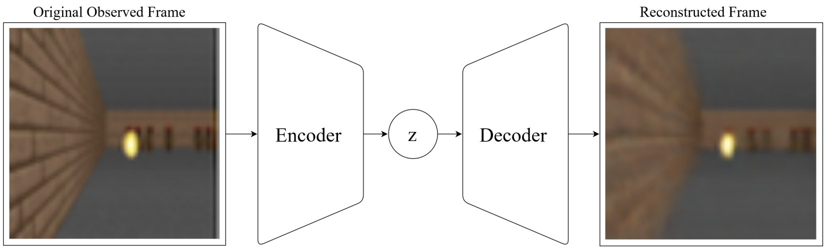
Computer vision is as large as image classification for aerial wilderness surveys, robotic path finding, obfuscation detection. One of the winners of the ICML 2018 best paper awards demonstrated that one of the main techniques for protecting image classifciation from adversarial examples could be easily circumvented.

Figure 1: Class-conditional samples generated by our model.

In another paper, the Google Deepmind team developed a GAN capable of generating high resolution synthetic images based on synthesizing ImageNet tags.



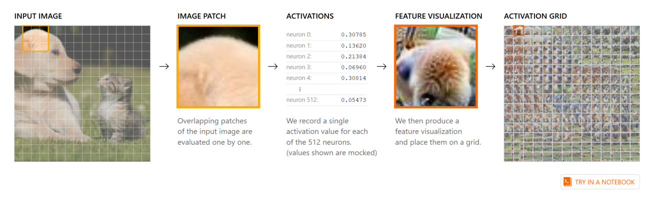
The winners of the CVPR 2018 Best Paper Award are using machine learning to try to disentangle the underlying visual structure of natural images. They identified relationships between 26 common visual tasks such as object recognition, depth estimation and edge detection in an attempt to learn what the structural similarities are between visual tasks that we find simple.



Flow diagram of a Variational Autoencoder. [31, 32]

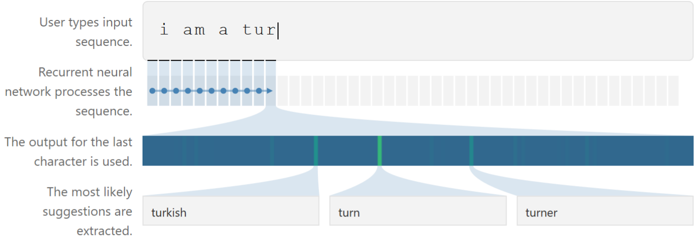
One of the other big paper to drop in 2018 was the World Models paper on data representation. In the paper and interactive website, they use deep encoding to model the environment around a reinforcement learner. This allows the training of an agent in a “virtual dream world” with an effectively infinite highly compressed amount of information, greatly speeding training.

Computer Vision and Image Analysis



Finally, one of the biggest recent developments in machine learning has been the creation of visual paradigms for opening the black box.

Computer Vision and Image Analysis



Autocomplete: An application that has a humanly interpretable output, while depending on both short and long-term contextual understanding. In this case, the network uses past information and understands the next word should be a country.

Finally, one of the biggest recent developments in machine learning has been the creation of visual paradigms for opening the black box. Algorithmic visualization is more than just a good advertisement for the field, good visualization allows us to understand and act upon the complicated algorithms we construct.

Big Ideas and Further Directions

There are many application and interesting topics that haven't even had time to cover in this survey:

- Time series and financial analysis.
- Instability Analysis
- Rare Event Prediction.
- Visualization.
- Ethics In Machine Learning.

Big Ideas and Further Directions

I hope this course has given you some sense of the world of machine learning, some fluency in the mathematics and computer science aspects, and some real hands on practical competency with data fitting.

Thank you for a great semester, I am excited to see how you all work to push the field forward, either as practitioner, enthusiasts, or researchers.