Supervised Learning:

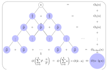
Classification in a nutshell

josh bowles[1,2]

[1]One on One Media
Software Developer

[2]Department of Philosophy (August 2012),
Experimental Epistemology and Machine Learning
University of Utah

Date / Bayes Group



Outline

This picture does not warm your heart

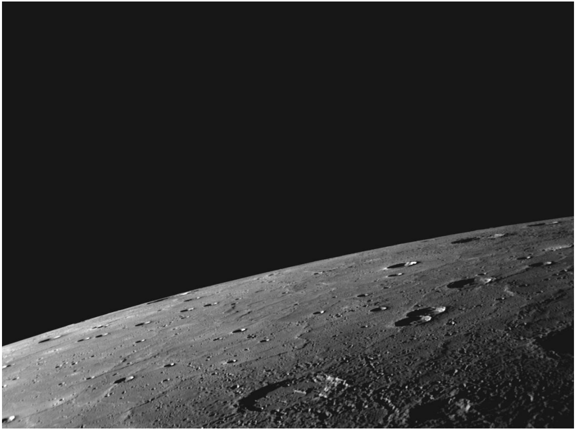


Figure: Mercury, NASA

Target (Approximation or Function)

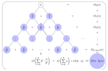
Less “science-fiction” name would be more appropriate: Target Approximation.

Machine Learning

Machine learning is really just a group of algorithms that all have similar end-game strategies: to make predictions given new data based on training data that specifies a target to be optimally approximated... or given the new data we want to (re)define a target function that will predict “probably approximately correct” (PAC) values

According to Ai taxonomy:

1. Fully observable
2. Stochastic
3. Continuous
4. Benign (non-adversarial)



Target (Approximation or Function)

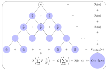
Less “science-fiction” name would be more appropriate: Target Approximation.

Machine Learning

Machine learning is really just a group of algorithms that all have similar end-game strategies: to make predictions given new data based on training data that specifies a target to be optimally approximated... or given the new data we want to (re)define a target function that will predict “probably approximately correct” (PAC) values

According to Ai taxonomy:

1. Fully observable
2. Stochastic
3. Continuous
4. Benign (non-adversarial)



Target (Approximation or Function)

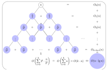
Less “science-fiction” name would be more appropriate: Target Approximation.

Machine Learning

Machine learning is really just a group of algorithms that all have similar end-game strategies: to make predictions given new data based on training data that specifies a target to be optimally approximated... or given the new data we want to (re)define a target function that will predict “probably approximately correct” (PAC) values

According to Ai taxonomy:

1. Fully observable
2. Stochastic
3. Continuous
4. Benign (non-adversarial)



Target (Approximation or Function)

Less “science-fiction” name would be more appropriate: Target Approximation.

Machine Learning

Machine learning is really just a group of algorithms that all have similar end-game strategies: to make predictions given new data based on training data that specifies a target to be optimally approximated... or given the new data we want to (re)define a target function that will predict “probably approximately correct” (PAC) values

According to Ai taxonomy:

1. Fully observable
2. Stochastic
3. Continuous
4. Benign (non-adversarial)

Target (Approximation or Function)

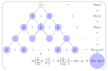
Less “science-fiction” name would be more appropriate: Target Approximation.

Machine Learning

Machine learning is really just a group of algorithms that all have similar end-game strategies: to make predictions given new data based on training data that specifies a target to be optimally approximated... or given the new data we want to (re)define a target function that will predict “probably approximately correct” (PAC) values

According to Ai taxonomy:

1. Fully observable
2. Stochastic
3. Continuous
4. Benign (non-adversarial)



This picture does not warm your heart either

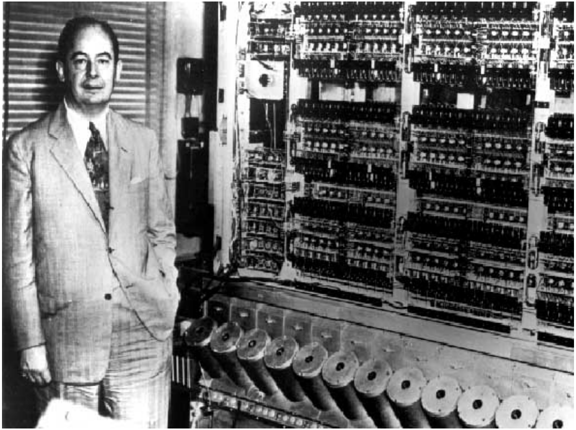


Figure: State of the Art Computer

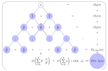
Functions that write functions

An idea I am exploring a little more (since this morning!)

Machine Learning and Meta-Programming

Machine Learning and Meta-Programming share some basics. Except ML attempts to design functions that write functions that perform better than their ancestors. Technical details depend on how we define “perform better”.

MP: Introspection (accessing language constructs at run-time) allows you to update/modify/etc. . . any construct at any time.
ML: Functions that need to update variables given unknown values of those variables, which in turn pipes output to another function. ML does this iteratively at massive scale.



Neural Networks

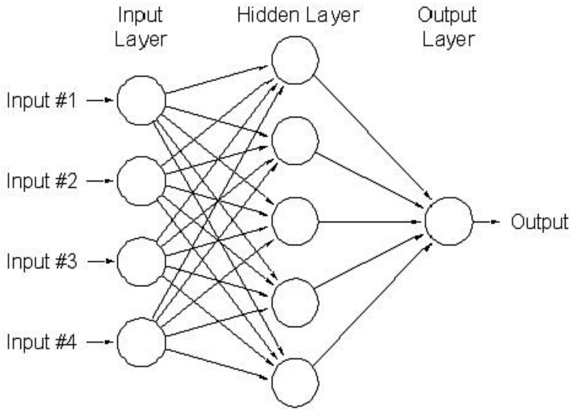
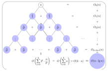


Figure: 3 Layer Neural Network, Hidden Layer is Meta-Programming Layer



Taxonomy

Learning via structural inference =

Not Deductive Logic

There is no contextual awareness and no domain-specific knowledge. There is no analogical learning (learning via analogy/example... what humans seem to be best at) that requires contextual-historical knowledge. This should not surprise anyone.

from Danks, to appear:4

Structural inference uses (relatively) domain-general algorithms whose success depends on the internal structure of the data, rather than features of the semantic content of the data

Danks, to appear. Learning. *Cambridge handbook to artificial intelligence*

Taxonomy

Learning via structural inference =

Not Deductive Logic

There is no contextual awareness and no domain-specific knowledge. There is no analogical learning (learning via analogy/example... what humans seem to be best at) that requires contextual-historical knowledge. This should not surprise anyone.

from Danks, to appear:4

Structural inference uses (relatively) domain-general algorithms whose success depends on the internal structure of the data, rather than features of the semantic content of the data

Danks, to appear. Learning. *Cambridge handbook to artificial intelligence*

More definitions

Supervised Learning algorithms assume that some variable X is designated as the target for prediction, explanation, or inference, and that the values of X in the dataset constitute the "ground truth" values for learning. That is, supervised learning algorithms use the known values of X to determine what should be learned.

More definitions...cont2

Tom M. Mitchell's widely cited definition

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

This picture is just freaky

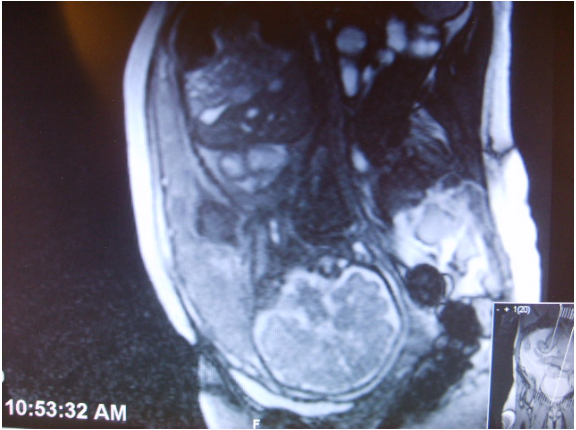


Figure: This has nothing to do with supervised learning

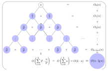
Outline

1. What is Supervised Learning?
    - Classification Tasks
2. What is Learning?
    - Targets
    - Training
    - In Practice
        - Keyword Classification

Binary Classification

Data: We need multiple sets of data; minimally, one for training and one for testing. A set of data is composed of, **at least**, the following:

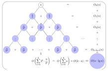
1. $\mathfrak{n}$ attributes: $[\mathcal{A}_0, \mathcal{A}_2, ..., \mathcal{A}_\mathfrak{n}]$; typically vectorized, such that $\mathfrak{n} \in \mathbb{R}$, or $\mathbb{R}^\mathfrak{n}$
2. a classification class $\mathbb{F}$ (= FINANCIAL)
3. a hypothesis space $\mathcal{H}$ (also called a class, but not to be confused with classification classes), for which each solution is a hypothesis $h \in \mathcal{H}$



Binary Classification

Data: We need multiple sets of data; minimally, one for training and one for testing. A set of data is composed of, **at least**, the following:

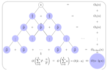
1. $\mathfrak{n}$ attributes: $[\mathcal{A}_0, \mathcal{A}_2, ..., \mathcal{A}_\mathfrak{n}]$; typically vectorized, such that $\mathfrak{n} \in \mathbb{R}$, or $\mathbb{R}^\mathfrak{n}$
2. a classification class $\mathbb{F}$ (= FINANCIAL)
3. a hypothesis space $\mathcal{H}$ (also called a class, but not to be confused with classification classes), for which each solution is a hypothesis $h \in \mathcal{H}$



Binary Classification

Data: We need multiple sets of data; minimally, one for training and one for testing. A set of data is composed of, **at least**, the following:

1. $\mathfrak{n}$ attributes: $[\mathcal{A}_0, \mathcal{A}_2, ..., \mathcal{A}_\mathfrak{n}]$; typically vectorized, such that $\mathfrak{n} \in \mathbb{R}$, or $\mathbb{R}^\mathfrak{n}$
2. a classification class $\mathbb{F}$ (= FINANCIAL)
3. a hypothesis space $\mathcal{H}$ (also called a class, but not to be confused with classification classes), for which each solution is a hypothesis $h \in \mathcal{H}$



Binary Classification cont...2

Goal: Define a classification model. Since this is binary we only need to define models with two values (positive and negative) for our class: $\mathfrak{p}, \mathfrak{n} \in \mathbb{F}$.
Our goal is: *hypothesis* $h \in \mathcal{H}$ that approximates $\mathbb{F}$ as closely as possible.

Supervision

All steps are being supervised by us... we are setting the values and determining both the goals and data at every point of the process, including our sense or intuition about how closely our hypothesis approximates the classification class

Binary Classification cont...3

1. data set = D
2. task = T
3. performance measure = M

Does our program “learn” from D to optimize task T given performance measure M? That is, does our program get better as approximating its target or re-defining its function to approximate correctly?
This is really general.

Binary Classification cont...3

1. data set = D
2. task = T
3. performance measure = M

Does our program "learn" from D to optimize task T given performance measure M? That is, does our program get better as approximating its target or re-defining its function to approximate correctly?
This is really general.

Binary Classification cont...3

1. data set = D
2. task = T
3. performance measure = M

Does our program “learn” from D to optimize task T given performance measure M? That is, does our program get better as approximating its target or re-defining its function to approximate correctly?
This is really general.

Continuous Classification

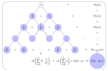
1. Get
2. To
3. This
4. Later

Continuous Classification

1. Get
2. To
3. This
4. Later

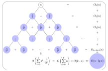
Continuous Classification

1. Get
2. To
3. This
4. Later



Continuous Classification

1. Get
2. To
3. This
4. Later



Google uses various multi-media classification

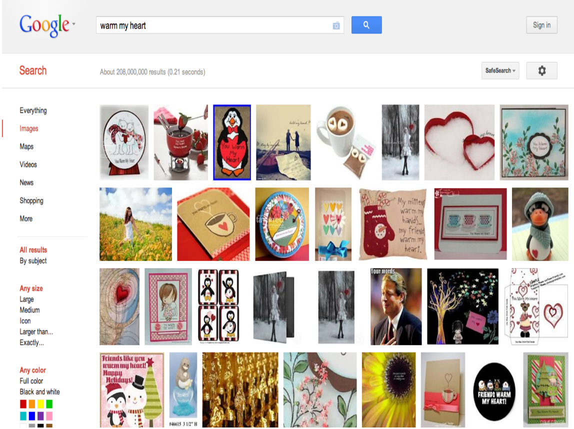
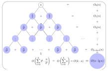


Figure: Text to visual classification uses various continuous classification model

Outline

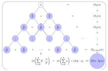
Learn towards targets

The most successful machine learning applications use **Math and Logic and Language** targets that are structural and not determined by context or domain-knowledge.



Outline

1. What is Supervised Learning?
   - Classification Tasks
2. What is Learning?
   - Targets
   - Training
   - In Practice
     - Keyword Classification



Learn from training

If

Outline

1. What is Supervised Learning?
   - Classification Tasks
2. What is Learning?
   - Targets
   - Training
   - In Practice
     - Keyword Classification

Learn from practice

It's not really about the machine learning anything... it is about YOU. The machine learns nothing, it's about us learning how to optimize tasks given the limits of our machines.

Note on Programming Languages

Ruby is great for rapid prototyping. If you want/need a scripting language, Python is still your best bet. Java seems to have the market share of machine learning.

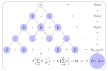
Sponsored Search

The Problem

Company has a large majority of its ad content and keyword bidding centered on financial aid incentives. It now needs to diversify its message content.

The Solution

Various Classifiers that will approximate the target classification Financial.



Sponsored Search

The Problem

Company has a large majority of its ad content and keyword bidding centered on financial aid incentives. It now needs to diversify its message content.

The Solution

Various Classifiers that will approximate the target classification Financial.

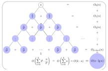
Classes

Classification Training

1. data_entities/financial.txt
2. data_entities/financial_not.txt

Data Training

1. data_training/
2. data_analysis/



Classes

Classification Training

1. data_entities/financial.txt
2. data_entities/financial_not.txt

Data Training

1. data_training/
2. data_analysis/

Classes

Classification Training

1. data_entities/financial.txt
2. data_entities/financial_not.txt

Data Training

1. data_training/
2. data_analysis/

Classes

Classification Training

1. data_entities/financial.txt
2. data_entities/financial_not.txt

Data Training

1. data_training/
2. data_analysis/