

Introduction to Machine Learning (CSCI-UA.0480-002)

**David Sontag
New York University**

Slides adapted from Luke Zettlemoyer, Pedro Domingos, and
Carlos Guestrin

Logistics

- **Class webpage:**
 - <http://cs.nyu.edu/~dsontag/courses/ml13/>
 - Sign up for mailing list!
- **Office hours:**
 - Tuesdays 5-6pm and by appointment.
 - 715 Broadway, 12th floor, Room 1204
- **Grader:** Chen-Chien Wang
 - Email: ccw352@nyu.edu

Evaluation

- About 7 homeworks (50%)
 - Both theory and programming
 - See collaboration policy on class webpage
- Midterm exam (25%)
- Project (20%)
- Course participation (5%)

Prerequisites

REQUIRED:

- **Basic algorithms** (CS 310)
 - Dynamic programming, algorithmic analysis
 - *Can be taken concurrently*

HELPFUL BUT NOT REQUIRED:

- **Linear algebra** (Math 140)
 - Matrices, vectors, systems of linear equations
 - Eigenvectors, matrix rank
 - Singular value decomposition
- **Multivariable calculus** (Math 123)
 - Derivatives, integration, tangent planes
 - Lagrange multipliers

Source Materials

No textbook required. Readings will come from freely available online material.

If you really want a book for an additional reference, these are good options:

- C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007
- K. Murphy, *Machine Learning: a Probabilistic Perspective*, MIT Press, 2012

A Few Quotes

- “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 the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, former Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, former CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, former CEO, Yahoo)

What is Machine Learning ?

(by examples)

Classification

from data to discrete classes

Spam filtering

data

prediction

☆ **Osman Khan** to Carlos [show details](#) Jan 7 (6 days ago) [Reply](#)

sounds good
+ok

Carlos Guestrin wrote:
Let's try to chat on Friday a little to coordinate and more on Sunday in person?

Carlos

Welcome to New Media Installation: Art that Learns

☆ **Carlos Guestrin** to 10615-announce, Osman, Michel [show details](#) 3:15 PM (8 hours ago) [Reply](#)

Hi everyone,

Welcome to New Media Installation:Art that Learns

The class will start tomorrow.
Make sure you attend the first class, even if you are on the Wait List.
The classes are held in Doherty Hall C316, and will be Tue, Thu 01:30-4:20 PM.

By now, you should be subscribed to our course mailing list: 10615-announce@cs.cmu.edu.
You can contact the instructors by emailing: 10615-instructors@cs.cmu.edu

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rlk [Spam](#) [X](#)

☆ **Jaquelyn Halley** to nherrlein, bcc: thehorney, bcc: anç [show details](#) 9:52 PM (1 hour ago) [Reply](#)

=== Natural WeightLOSS Solution ===

Vital Acai is a natural WeightLOSS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

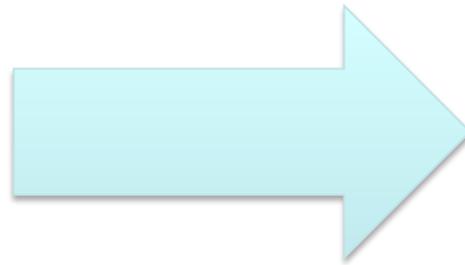
Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

- * Rapid WeightLOSS
- * Increased metabolism - BurnFat & calories easily!
- * Better Mood and Attitude
- * More Self Confidence
- * Cleanse and Detoxify Your Body
- * Much More Energy
- * BetterSexLife
- * A Natural Colon Cleanse



Spam
vs.
Not Spam

Weather prediction



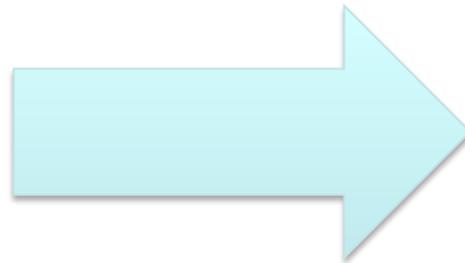
Regression

predicting a numeric value

Stock market



Weather prediction revisited



Temperature
72° F

Ranking

comparing items

Web search

The image shows a Google search interface. At the top left is the Google logo. A search bar contains the text "learning to rank". Below the search bar, a dropdown menu displays several suggestions: "learning to rank", "learning to rank for information retrieval" (with a link to "I'm Feeling Lucky »"), "learning to rank using gradient descent", and "learning to rank tutorial". To the right of the search bar is a blue search button with a magnifying glass icon. Below the search bar, the word "Search" is written in red. On the left side, there is a vertical navigation menu with options: "Web", "Images", "Maps", "Videos", "News", "Shopping", and "More". Below this menu, the location is set to "Manhattan, NY 10012" with a "Change location" link. At the bottom left, there is a "Show search tools" link. The main content area displays search results for "learning to rank". The first result is from Wikipedia, titled "Learning to rank - Wikipedia, the free encyclopedia", with the URL "en.wikipedia.org/wiki/Learning_to_rank". The snippet describes it as a type of supervised or semi-supervised machine learning problem. Below the snippet are links for "Applications", "Feature vectors", "Evaluation measures", and "Approaches". The second result is from Yahoo!, titled "Yahoo! Learning to Rank Challenge", with the URL "learningtorankchallenge.yahoo.com/". The snippet states that the challenge is closed and highlights some of the first ever Yahoo! results. The third result is a PDF titled "[PDF] Large Scale Learning to Rank" from Tufts University, with the URL "www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf". The snippet mentions pairwise learning to rank methods like RankSVM. The fourth result is from Microsoft Research, titled "Microsoft Learning to Rank Datasets - Microsoft Research", with the URL "research.microsoft.com/en-us/projects/mslr/". The snippet mentions two large scale datasets: L2R-WEB30k and L2R-WEB10K. The fifth result is also from Microsoft Research, titled "LETOR: A Benchmark Collection for Research on Learning to Rank ...", with the URL "research.microsoft.com/~letor/". The snippet describes the website as designed to facilitate research in Learning TO Rank (LETOR).

Google

learning to rank

learning to rank

learning to rank for information retrieval I'm Feeling Lucky »

learning to rank using gradient descent

learning to rank tutorial

Search

Web

Images

Maps

Videos

News

Shopping

More

Manhattan, NY 10012

Change location

Show search tools

[Learning to rank - Wikipedia, the free encyclopedia](https://en.wikipedia.org/wiki/Learning_to_rank)
en.wikipedia.org/wiki/Learning_to_rank
Learning to rank or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ...
[Applications](#) [Feature vectors](#) [Evaluation measures](#) [Approaches](#)

[Yahoo! Learning to Rank Challenge](https://learningtorankchallenge.yahoo.com/)
learningtorankchallenge.yahoo.com/
Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!

[\[PDF\] Large Scale Learning to Rank](https://www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf)
www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf
File Format: PDF/Adobe Acrobat - [Quick View](#)
by D Sculley - [Cited by 24](#) - [Related articles](#)
Pairwise **learning to rank** methods such as RankSVM give good performance, ... In this paper, we are concerned with **learning to rank** methods that can learn on ...

[Microsoft Learning to Rank Datasets - Microsoft Research](https://research.microsoft.com/en-us/projects/mslr/)
research.microsoft.com/en-us/projects/mslr/
We release two large scale datasets for research on **learning to rank**: L2R-WEB30k with more than 30000 queries and a random sampling of it L2R-WEB10K ...

[LETOR: A Benchmark Collection for Research on Learning to Rank ...](https://research.microsoft.com/~letor/)
research.microsoft.com/~letor/
This website is designed to facilitate research in **LEarning TO Rank** (LETOR). Much information about **learning to rank** can be found in the website, including ...

Given image, find similar images

 <p>1. Search mode: Theme 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>
 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>
 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>

 <p>1. Find similar by Theme 2. Search mode: Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>
 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>
 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>	 <p>1. Find similar by Theme OR 2. Find similar by Color / Texture</p>

Collaborative Filtering

Recommendation systems

The screenshot shows the Amazon.com interface with a navigation bar at the top. The main content area displays a list of four book recommendations under the heading 'Recommended for You'. Each recommendation includes a book cover, title, author, publication date, average customer review, and price information. The books are:

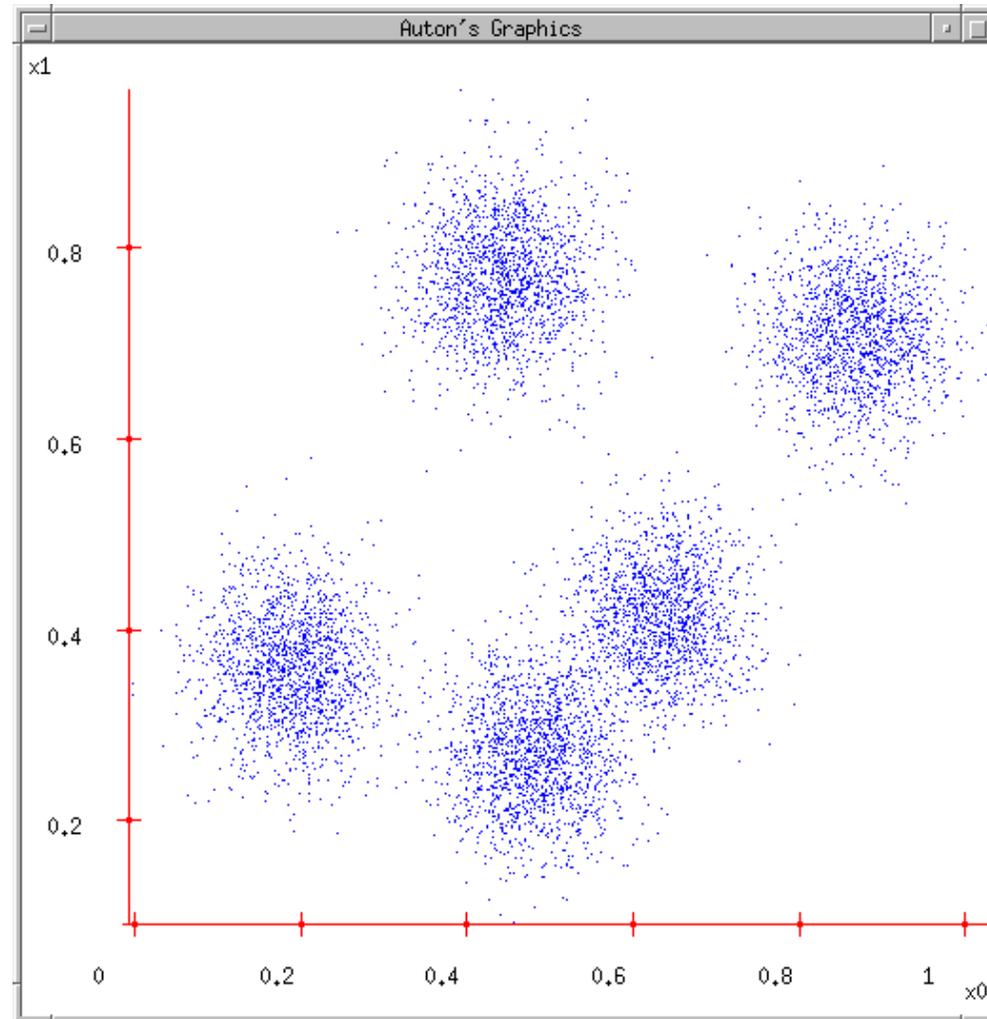
- Causality: Models, Reasoning and Inference** by Judea Pearl (September 14, 2009). Average Customer Review: 4.5 stars (10 reviews). List Price: \$50.00, Price: \$32.49. 61 used & new from \$28.00.
- The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century** by David Salsburg (May 1, 2002). Average Customer Review: 4.5 stars (26 reviews). List Price: \$18.99, Price: \$13.88. 81 used & new from \$9.00.
- The Eighth Day of Creation: Makers of the Revolution in Biology, 25th Anniversary Edition** by Horace Freeland Judson (November 1, 1996). Average Customer Review: 4.5 stars (10 reviews). List Price: \$56.00, Price: \$36.09. 59 used & new from \$26.95.
- The Machinery of Life** by David S. Goodsell (April 28, 2009). Average Customer Review: 4.5 stars (41 reviews). List Price: \$25.00, Price: \$17.49. 92 used & new from \$12.00.

Each recommendation also includes an 'Add to Cart' and 'Add to Wish List' button, and a 'Rate this item' section with a star rating and a 'Fix this' link.

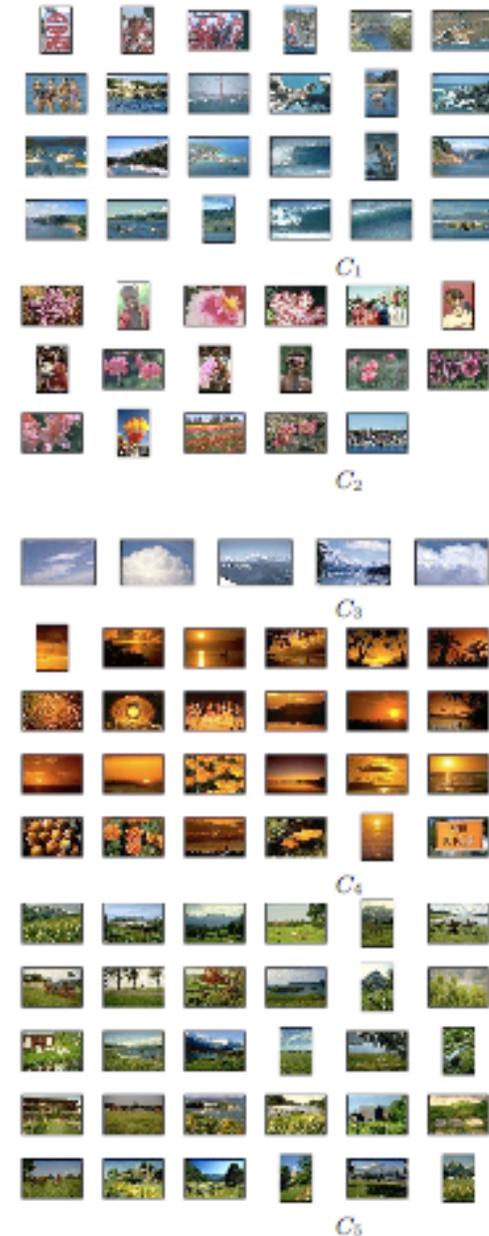
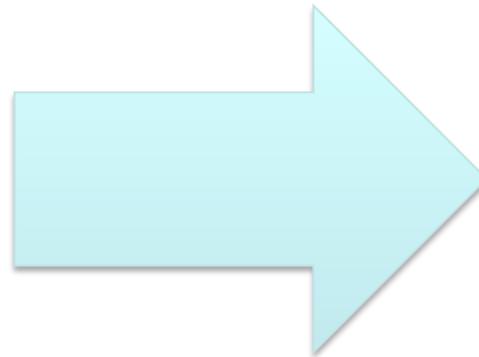
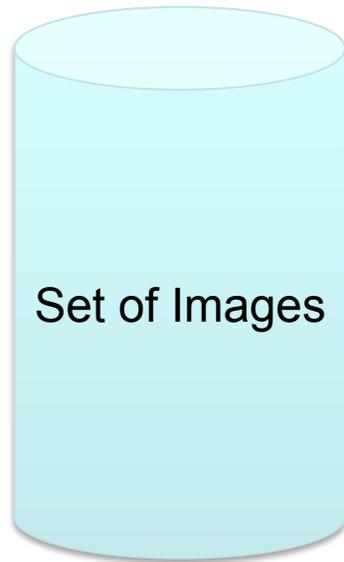
Clustering

discovering structure in data

Clustering Data: Group similar things



Clustering images



[Goldberger et al.]

Clustering web search results

Clusty

web news images wikipedia blogs jobs more »

race Search advanced preferences

clusters sources sites

All Results (238) remix

- Car (28)
 - Race cars (7)
 - Photos, Races Scheduled (5)
 - Game (4)
 - Track (3)
 - Nascar (2)
 - Equipment And Safety (2)
 - Other Topics (7)
- Photos (22)
- Game (14)
- Definition (13)
- Team (18)
- Human (8)**
 - Classification Of Human (2)
 - Statement, Evolved (2)
 - Other Topics (4)
- Weekend (8)
- Ethnicity And Race (7)
- Race for the Cure (8)
- Race Information (8)

more | all clusters

find in clusters: Find

Cluster Human contains 8 documents.

Search Results

- [Race \(classification of human beings\) - Wikipedia, the free ...](#)   

The term **race** or racial group usually refers to the concept of dividing **humans** into populations or groups on the basis of various sets of characteristics. The most widely used **human** racial categories are based on visible traits (especially skin color, cranial or facial features and hair texture), and self-identification. Conceptions of **race**, as well as specific ways of grouping **races**, vary by culture and over time, and are often controversial for scientific as well as social and political reasons. History · Modern debates · Political and ...
[en.wikipedia.org/wiki/Race_\(classification_of_human_beings\)](http://en.wikipedia.org/wiki/Race_(classification_of_human_beings)) - [cache] - Live, Ask
- [Race - Wikipedia, the free encyclopedia](#)   

General. **Racing** competitions The **Race** (yachting **race**), or La course du millénaire, a no-rules round-the-world sailing event; **Race** (biology), classification of flora and fauna; **Race** (classification of **human** beings) **Race** and ethnicity in the United States Census, official definitions of "**race**" used by the US Census Bureau; **Race** and genetics, notion of racial classifications based on genetics. Historical definitions of **race**; **Race** (bearing), the inner and outer rings of a rolling-element bearing. **RACE** in molecular biology "Rapid ... General · Surnames · Television · Music · Literature · Video games
en.wikipedia.org/wiki/Race - [cache] - Live, Ask
- [Publications | Human Rights Watch](#)   

The use of torture, unlawful rendition, secret prisons, unfair trials, ... Risks to Migrants, Refugees, and Asylum Seekers in Egypt and Israel ... In the run-up to the Beijing Olympics in August 2008, ...
www.hrw.org/background/usa/race - [cache] - Ask
- [Amazon.com: Race: The Reality Of Human Differences: Vincent Sarich ...](#)   

Amazon.com: **Race: The Reality Of Human Differences: Vincent Sarich, Frank Miele: Books ...** From Publishers Weekly Sarich, a Berkeley emeritus anthropologist, and Miele, an editor ...
www.amazon.com/Race-Reality-Differences-Vincent-Sarich/dp/0813340861 - [cache] - Live
- [AAPA Statement on Biological Aspects of Race](#)   

AAPA Statement on Biological Aspects of **Race** ... Published in the American Journal of Physical Anthropology, vol. 101, pp 569-570, 1996 ... PREAMBLE As scientists who study **human** evolution and variation, ...
www.physanth.org/positions/race.html - [cache] - Ask
- [race: Definition from Answers.com](#)   

race n. A local geographic or global **human** population distinguished as a more or less distinct group by genetically transmitted physical
www.answers.com/topic/race-1 - [cache] - Live
- [Dopefish.com](#)   

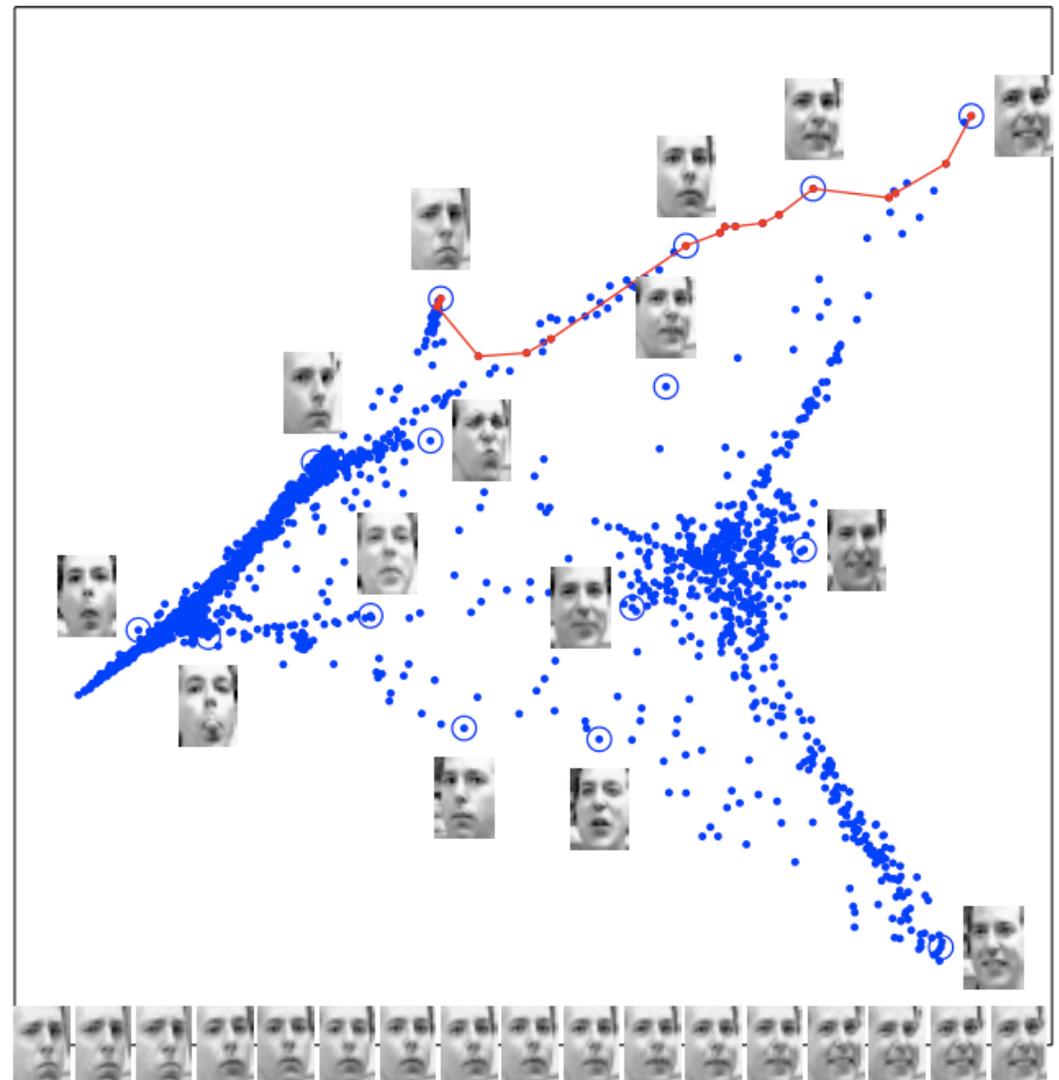
Site for newbies as well as experienced Dopefish followers, chronicling the birth of the Dopefish, its numerous appearances in several computer games, and its eventual take-over of the **human race**. Maintained by Mr. Dopefish himself, Joe Siegler of Apogee Software.
www.dopefish.com - [cache] - Open Directory

Embedding

visualizing data

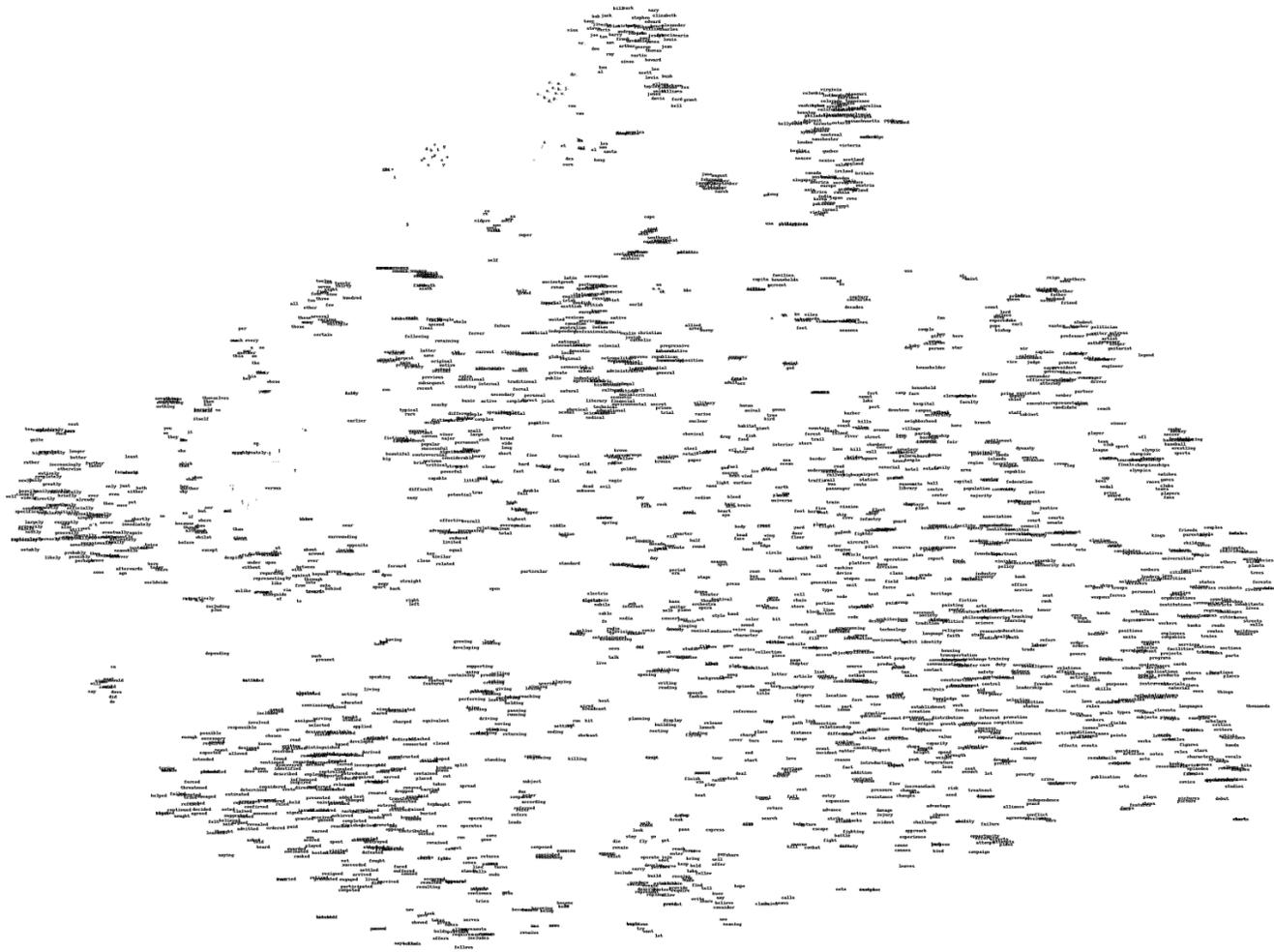
Embedding images

- Images have thousands or millions of pixels.
- Can we give each image a coordinate, such that similar images are near each other?



[Saul & Roweis '03]

Embedding words

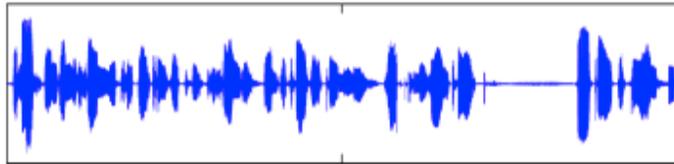


[Joseph Turian]

Structured prediction

from data to discrete classes

Speech recognition



Natural language processing

I need to hide a body
noun, verb, preposition, ...



Growth of Machine Learning

- Machine learning is preferred approach to
 - Speech recognition, Natural language processing
 - Computer vision
 - Medical outcomes analysis
 - Robot control
 - Computational biology
 - Sensor networks
 - ...
- This trend is accelerating
 - Big data
 - Improved machine learning algorithms
 - Faster computers
 - Good open-source software

Supervised Learning: find f

- **Given:** Training set $\{(x_i, y_i) \mid i = 1 \dots N\}$
- **Find:** A good approximation to $f : X \rightarrow Y$

Examples: what are X and Y ?

- **Spam Detection**
 - Map email to {Spam, Not Spam}
- **Digit recognition**
 - Map pixels to {0,1,2,3,4,5,6,7,8,9}
- **Stock Prediction**
 - Map new, historic prices, etc. to \mathfrak{R} (the real numbers)

A Supervised Learning Problem

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

- Our goal is to find a function $f : X \rightarrow Y$
 - $X = \{0,1\}^4$
 - $Y = \{0,1\}$
- **Question 1:** How should we pick the *hypothesis space*, the set of possible functions f ?
- **Question 2:** How do we find the best f in the hypothesis space?

Most General Hypothesis Space

Consider all possible boolean functions over four input features!

- 2^{16} possible hypotheses
- 2^9 are consistent with our dataset
- How do we choose the best one?

x_1	x_2	x_3	x_4	y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	?
1	0	0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	?
1	1	0	0	0
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

A Restricted Hypothesis Space

Consider all conjunctive boolean functions.

- 16 possible hypotheses

- None are consistent with our dataset

- How do we choose the best one?

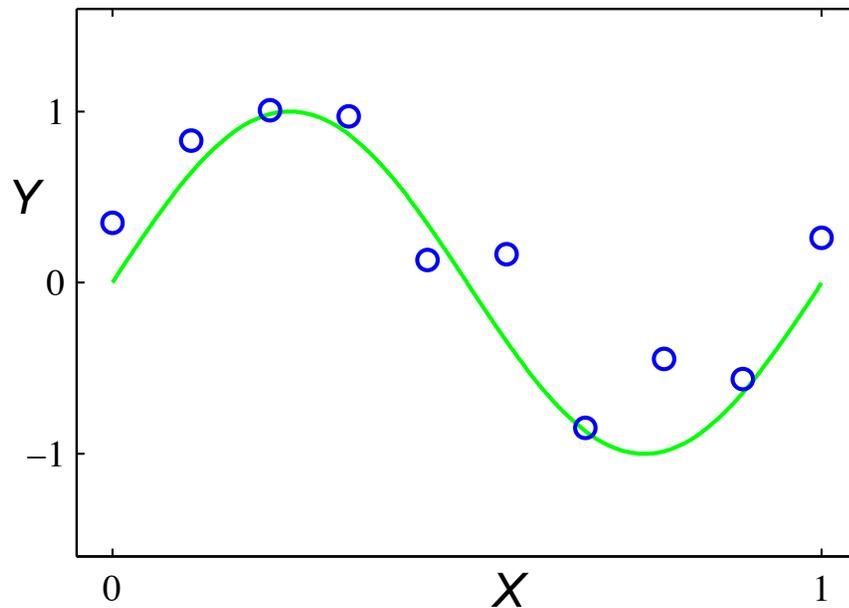
Rule	Counterexample
$\Rightarrow y$	1
$x_1 \Rightarrow y$	3
$x_2 \Rightarrow y$	2
$x_3 \Rightarrow y$	1
$x_4 \Rightarrow y$	7
$x_1 \wedge x_2 \Rightarrow y$	3
$x_1 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_4 \Rightarrow y$	3
$x_2 \wedge x_3 \Rightarrow y$	3
$x_2 \wedge x_4 \Rightarrow y$	3
$x_3 \wedge x_4 \Rightarrow y$	4
$x_1 \wedge x_2 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_4 \Rightarrow y$	3
$x_1 \wedge x_3 \wedge x_4 \Rightarrow y$	3
$x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

Second example: Regression

Dataset: 10 (X,Y) points generated from a sin function, with noise



- Regression:

- $f : X \rightarrow Y$

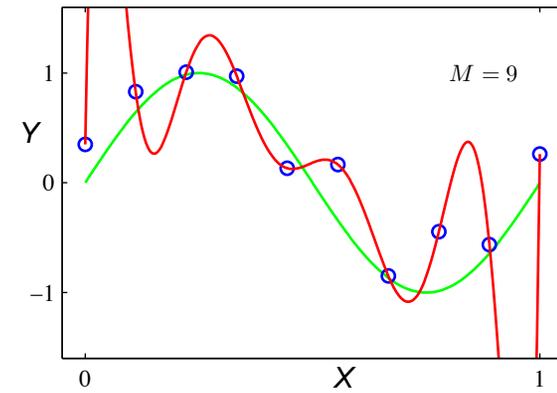
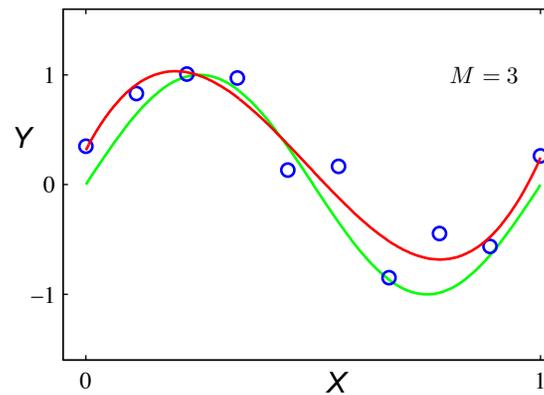
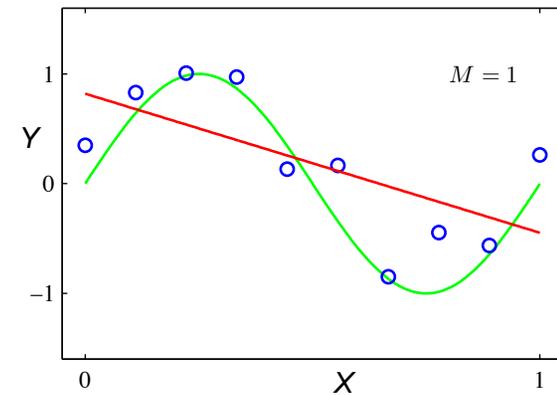
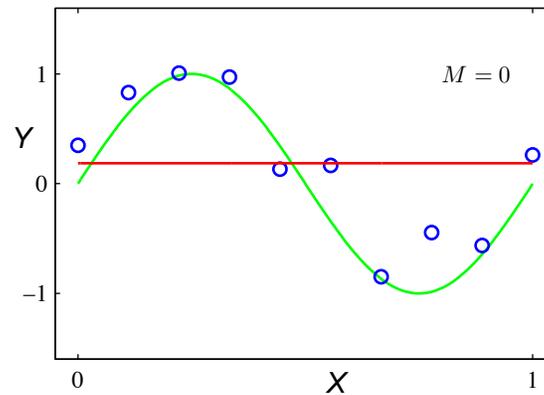
- $X = \mathfrak{R}$

- $Y = \mathfrak{R}$

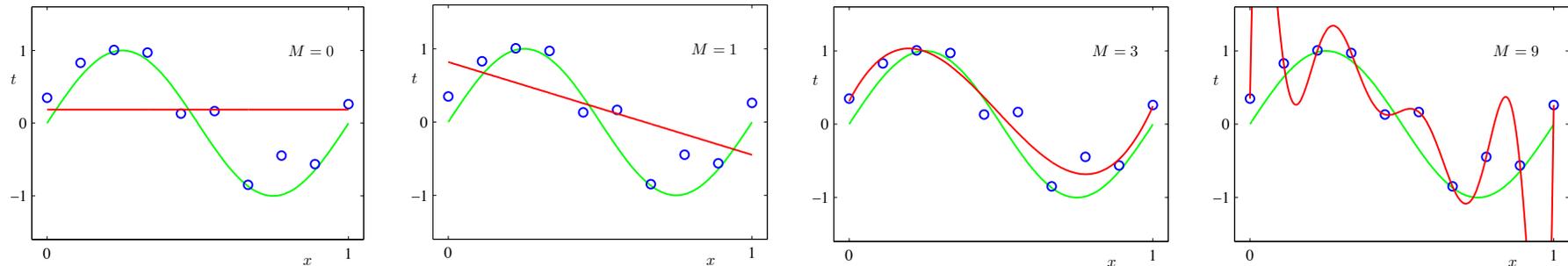
Degree-M Polynomials

How about letting f be a degree M polynomial?

•Which one is **best**?



Hypo. Space: Degree-N Polynomials



We measure error using a *loss function* $L(y, \hat{y})$

For regression, a common choice is squared loss:

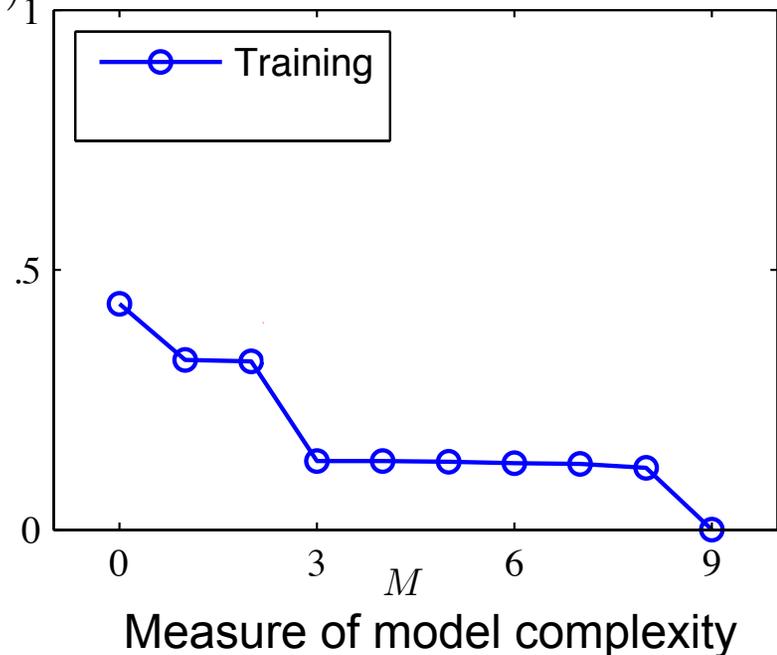
$$L(y_i, f(x_i)) = (y_i - f(x_i))^2$$

Squared error

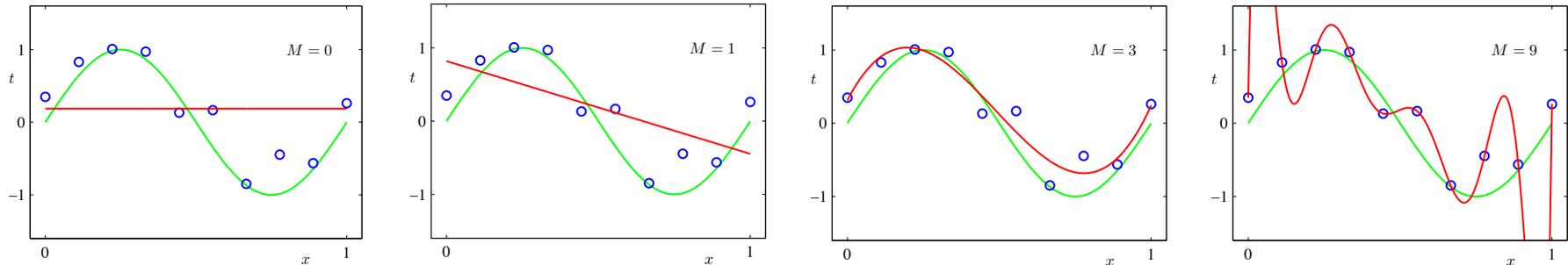
The *empirical loss* of the function f applied to the training data is then:

$$\frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2$$

Learning curve



Hypo. Space: Degree-N Polynomials



We measure error using a *loss function* $L(y, \hat{y})$

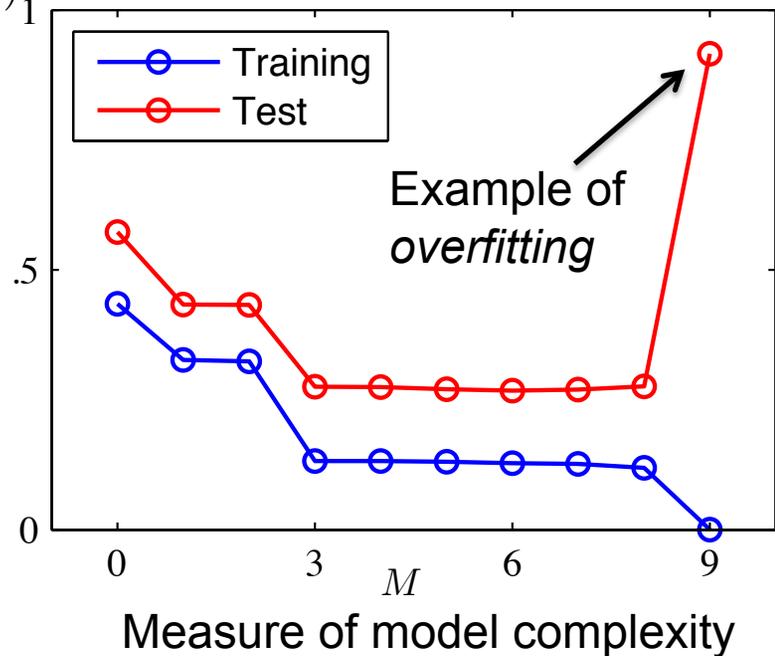
For regression, a common choice is squared loss:

$$L(y_i, f(x_i)) = (y_i - f(x_i))^2 \quad \text{Squared error}$$

The *empirical loss* of the function f applied to the training data is then:

$$\frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2$$

Learning curve



Occam's Razor Principle

- William of **Occam**: Monk living in the 14th century
- Principle of parsimony:

“One should not increase, beyond what is necessary, the number of entities required to explain anything”

- When **many** solutions are available for a given problem, we should select the **simplest** one
- But what do we mean by **simple**?
- We will use **prior knowledge** of the problem to solve to define what is a simple solution

Example of a prior: smoothness

[Samy Bengio]

Key Issues in Machine Learning

- How do we choose a hypothesis space?
 - Often we use **prior knowledge** to guide this choice
- How can we gauge the accuracy of a hypothesis on unseen data?
 - **Occam's razor**: use the *simplest* hypothesis consistent with data! This will help us avoid overfitting.
 - **Learning theory** will help us quantify our ability to **generalize** as a function of the amount of training data and the hypothesis space
- How do we find the best hypothesis?
 - This is an **algorithmic** question, the main topic of computer science
- How to model applications as machine learning problems?
(engineering challenge)