Supervised Learning

Dr. Jianlin Cheng

Electrical Engineering and Computer Science Department University of Missouri, Columbia Fall, 2019

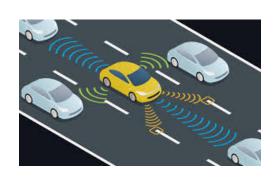
Syllabus

- Course web site: http://calla.rnet.missouri.edu/cheng_co
 - http://calla.rnet.missouri.edu/cheng_courses/supervised_learning/index.htm
- Location: Lafferre Hall E3403; Time:
 MWF 11:00 pm 11:50 am
- Assignments: 4 homework assignments, one group project (up to 4 students)
- **Grading**: participation (25%), assignment (30%), project report (20%), project representation (25%); grade scale (A+, A, A-, B+, B, B-, C+, C, C-, and F)
- Questions / Assignment submission: <u>mumachinelearning@gmail.com</u>

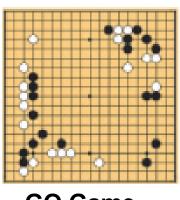
Topics

- Introduction to machine learning and Bayes optimal learning rule
- Learning distributions, parametric distribution, Maximum Likelihood Estimation (MLE) and Maximum a Posterior Estimation (MAP)
- MLE, MAP, Bayes Optimal Classifier, Naïve Bayes Classifier, Generative Classifier
- Discriminative classifier and logistic regression
- Linear and non-linear regression
- Nonparametric methods for density estimation, classification and regression
- Model selection
- Boosting
- Deep learning (a focus)

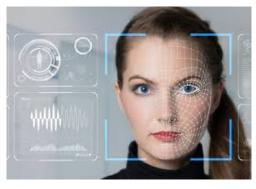
Artificial Intelligence (AI) Revolution



Self Driving Car



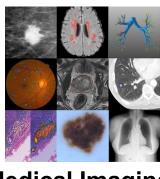
GO Game



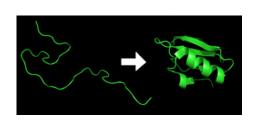
Face Recognition



Voice Recognition



Medical Imaging

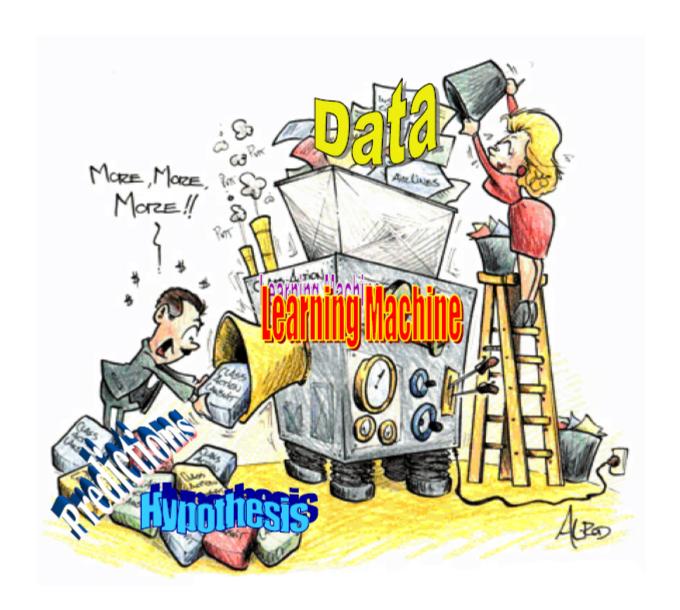


Protein Folding

What is Artificial Intelligence (AI) and Deep Learning?

- <u>AI</u>: a technology to enable computational devices to gain human intelligence (speak, walk, drive, think, ..., <u>learning</u>)
- Machine Learning: an area of AI that enables computer to learn from data to acquire any specific intelligence
- <u>Deep Learning</u>: the most powerful machine learning technology mimicking human learning

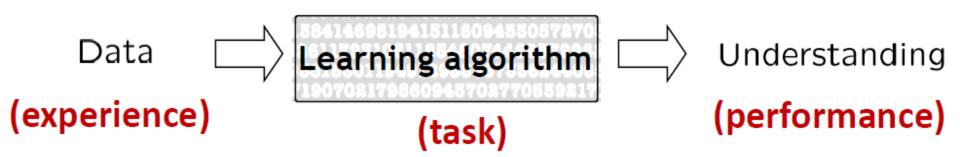
What is Machine Learning?



What is Machine Learning?

Study of algorithms that

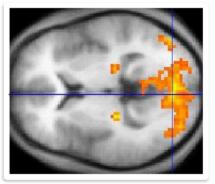
- improve their <u>performance</u>
- at some <u>task</u>
- with <u>experience</u>



Question

 What is the difference between a machine learning and a traditional instruction-based program such as a calculator, windows, Linux?

Decoding thoughts from brain scans



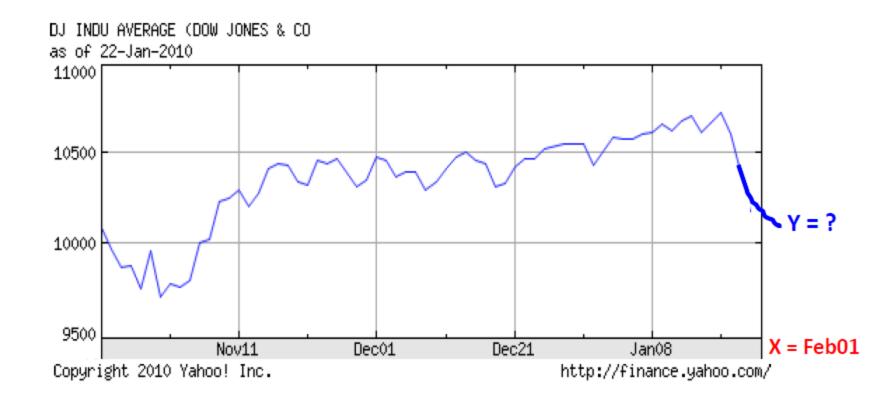


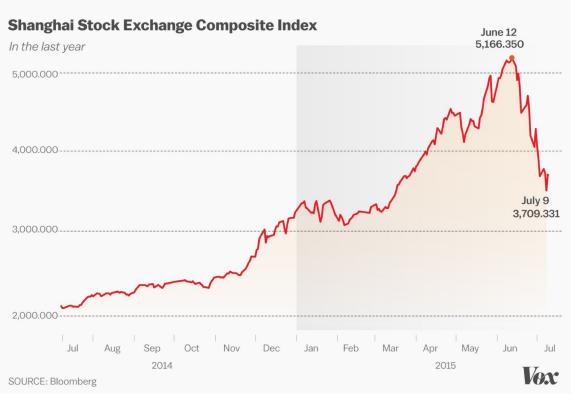
Rob a bank ...





Stock Market Prediction







Dow Jones August 12-16, 2019

YTD

1 year

Market Summary > Dow Jones Industrial Average INDEXDJX: .DJI

1 month

6 months



Max

5 years

25,886.01 +306.62 (1.20%)

Aug 16, 4:51 PM EDT · Disclaimer

5 days

1 day



Document classification



Spam filtering

Welcome to New Media Installation: Art that Learns

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.

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only \$5.95 for shipping mfw rlk | Spam | X



Spam/ Not spam

D

=== Natural WeightL0SS 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 WeightL0SS
- * Increased metabolism BurnFat & calories easily!
- * Bottor Mood and Attitude

Cars navigating on their own



Boss, the self-driving SUV 1st place in the DARPA Urban Challenge.

Photo courtesy of Tartan Racing.



Google Self-Driving Car

Laser -

This sensor gives the vehicle a 360-degree understanding of its environment so the car can sense objects in front of, beside, and behind itself at the same time, all the time. The laser also helps the vehicle to determine its location in the world.

Processor

Information from the sensors is cross-checked and processed by the software so that different ~ objects around the vehicle can be sensed and differentiated accurately, and safe driving decisions can then be made based on all the information received.

Position sensor

This sensor, located in the wheel hub, detects the rotations made by the wheels of the car to help the vehicle understand its position in the world.

Similar to the way a person's inner ear gives them a sense of motion and balance, this sensor, located in the interior of the car, works to give the car a clear sense of orientation.

Safety drivers

Radar

Drivers also test the vehicles daily, reporting feedback on how to make the ride more safe and comfortable.

Orientation sensor

This sensor detects vehicles far ahead and measures their speed so that the car can safely slow down or speed up with other vehicles on the road.

- The best helicopter pilot is now a computer!
 - it runs a program that learns how to fly and make acrobatic maneuvers by itself!
 - no taped instructions, joysticks, or things like that ...





Drones



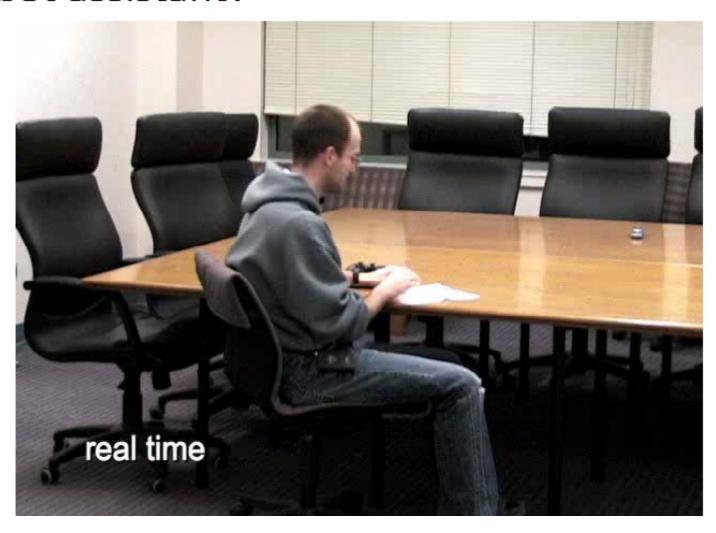


Amazon

US Air Force

Robot assistant?

[http://stair.stanford.edu/]



Industrial Robots



Natural language processing and speech recognition

 Now most pocket Speech Recognizers or Translators are running on some sort of learning device --- the more you play/use them, the smarter they become!

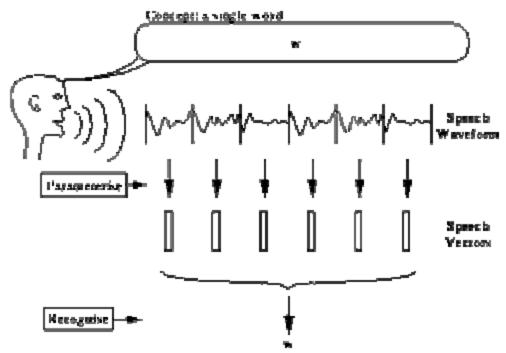


Fig. 1.2 Isolated Word Problem



Siri



 Behind a security camera, most likely there is a computer that is learning and/or checking!







Face Recognition









Face Recognition







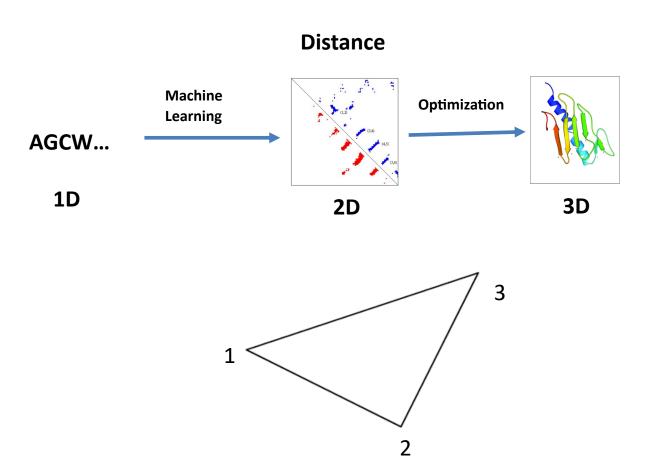


Scientific Discovery

Which genetic mutation is responsible for breast cancer and supplied to the su

gtacttgaggaattggegategtagegattteeceegeegtaaagtteetgateetegttgtttttgtacateataaagteeggattetgetegtegeegaagatgggaaegaagetgeeaaagetgaaggtetgettgaggtgetgg :cccagctggataaccttgctgtacagatcggcatctgcctggagggcacgatcgaaatccttccagtggacgaacttcacctgctcgctgggaatagcgttgttgtcaagcagctcaaggagcgtattcgagttgacgggctgcacca tgegteetatgegaggaatggtetttaggttetttatggeaaagttetegeetegettgeeeageeggtaegttettggtgatetttaggaagaateetggaetaetgte igactyttatgatteteatgetgatgegaetgaagetteacetgaeteetgeteeacaattygtggeetttatatagegagateeaceegeate<mark>t</mark>tgegtggaatagaaa <mark>t</mark>gegggtgaeteeaggaattageattategateggaaagi icteaaatagetgegateaetegggageagggtaaagtegeetegaaacaggaagetgaageatettetataaatacaeteaaagegategtteegaggegagtetggtta .categetgegttteggeagetaattgeettttagaaattatttteeeatttegagaaaetegtgtgggatgeeggatgeggettteaa/caettetggeeegggateggdttggggteaeattgtetgegggetetattgtetegatee taaccaaaaaacagcagataaacaaaaacggcaaagaaaattgccacagagttgtcacgctttgttgcacaaacatttgtgcagaaAagtgaaaagcttttagccattattaagctttttcctcagctcgctggcagcacttgcgaatg gatgtteeteataaatgaaaattaatgttigetetaegeteeaeegaaetegetigtiiggiggatiggetggetaategeggetagateeeaggeggtaiaaeeetiitegetteateagiigigaaaeeagaiggeiggigtiitigg .geggaetecectegaaegetetegaaateaagtggetttecageeggeeegetgggeegetegeeeactggaeeggtatteeegggeeacageeggtaceageeggtaceeggeeacggeeacactgtaeeggeaeeggtataateetegeeagaeteggegetgataaggeeeaatg tecgcaggcgtctattttatgccaaggaccgttcttcttcagctttcggctcgagtatttgttgttgtgccatgttggttacgatggcaatcgcggttacagttatgcaaatgcaggcgaataccgctcactgacaatgaacggcgtcttgttteatgetgacatteatatteatteetttggttttttgtettegaeggaetgaaaagtgeggagaaaeeeaaaaeaga/gegegeaaagegeegttaatatgegaaeteattgaagttateaeaaeaeatateea gagtaaacgaggegtatgegeaacetgeacetggeggaegeggegtatgegeaatgtgeaattegettacettetegtt/gegggteaggaacteecagatgggaatggee/atgaegagetgatetgaatgtggaaggegeecageag gaktacktkegcegcagkegtegtegtegttgetgettkkatgtkgegkackeegcackacaeggagagkkeaggggaktegtgeteegtgaketgtgteegtgteegtgggteaaktgeaeggkkeggtkgtgtgkacekkegtgttttttttttagggeeeaataaaagegettttgtggeggettgatagattateaettggttteggtggetageeaagVggetttettetgteegaegeaettaattgaatt egtgtgteteagettgaaaegeaaaagettgttteaeaeateggttteteggeaagatggggggteagteggtetal Wilneggggggegeateggategteggteggegaatteeaageggaagegaageggaagegaageggaagegaageggaagggaageggaageggaageggaageggaageggaageggaageggaageggaaggaagggaaggaaggaagggaagggaaggaggaaggaggaaggaaggaggaaggaggaaggaggaagggaggaggaggaggaggag .geggeagegggtggggategggagececeegaaaaaaacagaggegeaegtegatgeeateggggaattggaaceteaatgtgtgggaatgtttaaatattetgtgttaggtagtgtagttteatagaetatagatteteataeaga gteettegageegattataeaegaeageaaaatattteagtegegettgggeaaaaggettaageaegaeteeeagteeeeettaeatttgtetteetaageeeetggageeaetateaaaettgttetaegettgeaetgaaaata; :caaagtaaacaatcaaaaagaccaaaaaacaataacaaccagcaccgagtcgaacatcagtgaggcattgcaaaaatttcaaagttaagttggtcgtcgtcatcgcgtctgagtccgatcaagccgggcttgtaattgaagttgttgatgi actggattgtggggaattetggteageataettaacageageeegetaattaageaaaataacatateeaaatteeagaatgegaeggegeeateateetgtttgggaatteaatteggggeagategtttaatteaatteaaaggt .aagggagcagaagaatgcgatcgctggaatttcctaacatcacggaccccataaatttgataagcccgagctcgcttgcgttgagtcagcccccacatccccacatccccgccaaaatccccgccaaaagaagacagctgggttgttgactcgccagat elegan and to attempt of the later of the la tte w C t t tgansa act C thec G gt was confirmed to the startest at tect the tent at case to case to gtacttgaggaattggcgatcgtagcgatttccccccgccgta 🛁 ttcctgatcctcgttgtttttgtacatcataaagtccggattctgctcgtcgccgaagatgggaacgaagctgccaaagctgcaaagctgctgcttgaggtgctgg :cccagctggataaccttgctgtacagatcggcatctgcctggagggcacgatcgaaatccttccagtggacgaacttcacctgctcgctgggaatagcgttgttgtcaagcagctcaaggagcgtattcgagttgacgggctgcacca

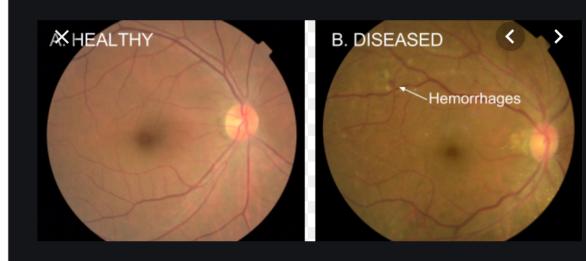
Protein Folding



Google's AlphaFold (1st) and MU's MULTICOM (3rd)

Disease Diagnosis

https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html









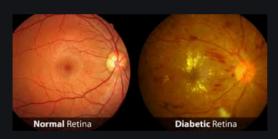


Google Al Blog: Deep Learning for Detection of Diabetic Eye Disease

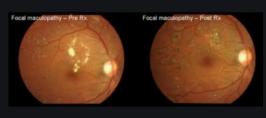
Images may be subject to copyright. Learn More

Related images

See more



nv | Eye Care | Regular eye exam...
nvmyeyes.com



The Retina Clinic - About your ret... theretinaclinic.com

Question

- In your opinion, what other problems can be addressed by machine learning?
- Why?

Big Data Challenges & Machine Learning

VOLUME

- Terabytes
- Records
- Transactions
- Tables, files

3 Vs of Big Data

- Batch
- Near time
- Real time
- Streams

- Structured
- Unstructured
- Semistructured
- All the above

VELOCITY

VARIETY

What this course is about

- Covers a wide range of Machine Learning techniques
 - from basic to state-of-the-art
- You will learn about the methods you heard about:
 - Naïve Bayes, logistic regression, nearest-neighbor, decision trees, boosting, neural nets, overfitting, regularization, dimensionality reduction, PCA, error bounds, VC dimension, SVMs, kernels, margin bounds, K-means, EM, mixture models, semi-supervised learning, HMMs, graphical models, active learning, reinforcement learning...

 Deep Learning
- Covers algorithms, theory and applications
- It's going to be fun and hard work ⁽²⁾

Machine Learning Tasks

Broad categories -

Supervised learning

Classification, Regression

Unsupervised learning

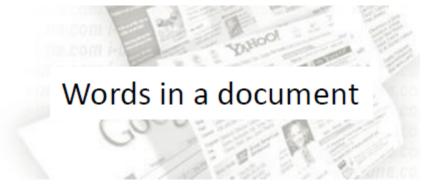
Density estimation, Clustering, Dimensionality reduction

- Semi-supervised learning
- Active learning
- Reinforcement learning
- Many more ...

Supervised Learning



Label Space ${\mathcal Y}$





"Sports" "News" "Science"

...

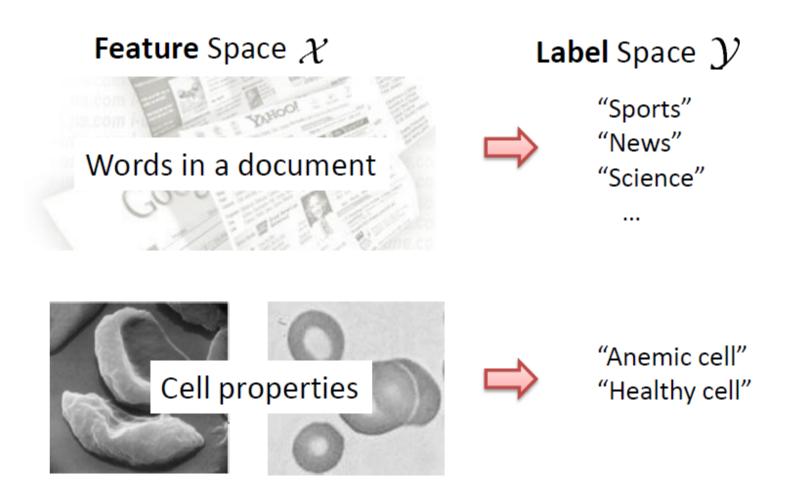




Share Price "\$ 24.50"

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Supervised Learning - Classification

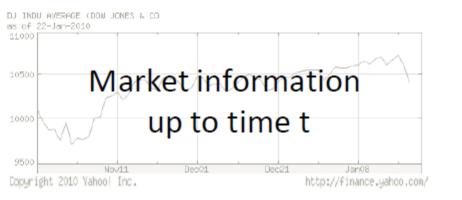


Discrete Labels

Supervised Learning - Regression

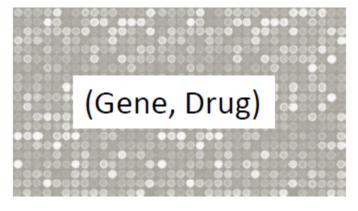
Feature Space \mathcal{X}

Label Space ${\mathcal Y}$





Share Price "\$ 24.50"

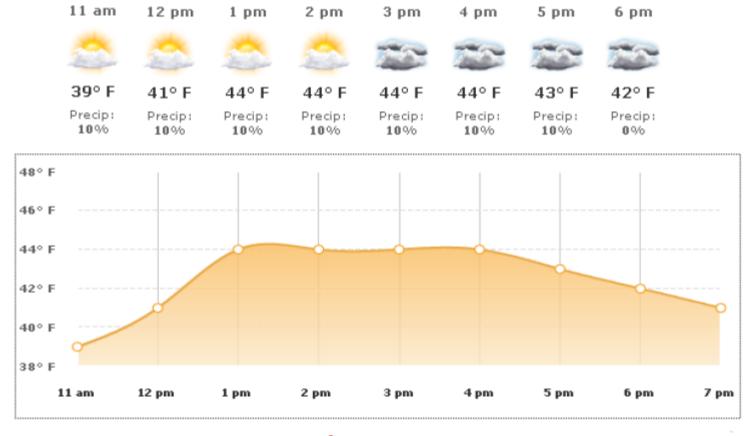




Expression level "0.01"

Supervised Learning problems

Features? Labels? Classification/Regression?



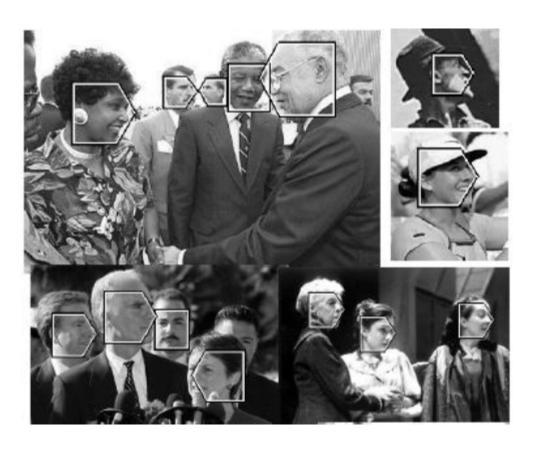
Temperature/Weather prediction

Supervised Learning problems

Features?

Labels?

Classification/Regression?



Face Detection

Supervised Learning problems

Features?

Labels?

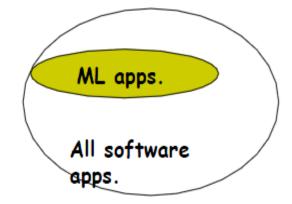
Classification/Regression?



Environmental Mapping

Growth of Machine Learning

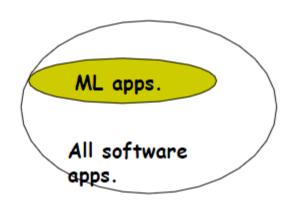
- Machine learning already the preferred approach to
 - Speech recognition, Natural language processing
 - Computer vision
 - Medical outcomes analysis
 - Robot control
 - ...



This ML niche is growing (why?)

Growth of Machine Learning

- Machine learning already the preferred approach to
 - Speech recognition, Natural language processing
 - Computer vision
 - Medical outcomes analysis
 - Robot control
 - ...



- This ML niche is growing
 - Improved machine learning algorithms
 - Increased data capture, networking
 - Software too complex to write by hand
 - New sensors / IO devices
 - Demand for self-customization to user, environment

Function Approximation

Setting:

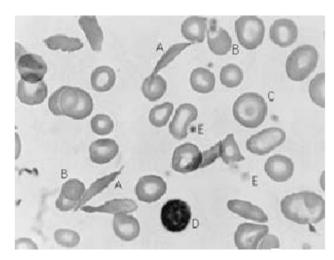
- Set of possible instances X
- Unknown target function f: X→Y
- Set of function hypotheses H={ h | h: X→Y }

Given:

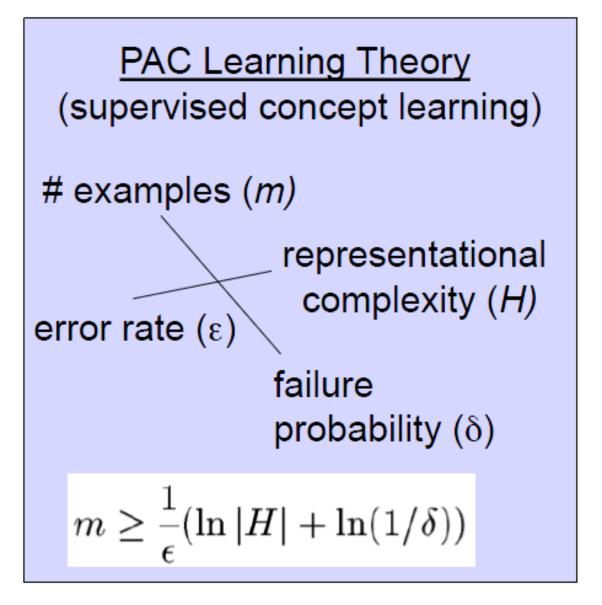
Training examples {<x_i,y_i>} of unknown target function f

Determine:

Hypothesis h ∈ H that best approximates f



Probably Approximately Correct Learning



Occam's Razor – When everything is equal, a simple solution is better.

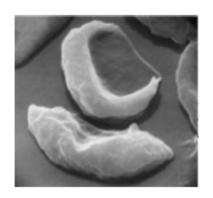


Supervised Learning Task

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

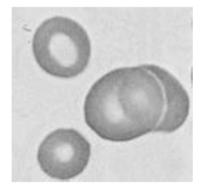
X - test data

 \equiv Construct **prediction rule** $f: \mathcal{X} \rightarrow \mathcal{Y}$





"Anemic cell (0)"

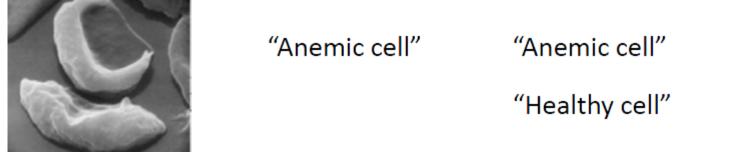




"Healthy cell (1)"

Performance:

loss(Y, f(X)) - Measure of closeness between true label Y and prediction f(X)



 $loss(Y, f(X)) = 1_{\{f(X) \neq Y\}}$ 0/1 loss

0

1

Performance:

loss(Y, f(X)) - Measure of closeness between true label Y and prediction f(X)

X	Share price, Y	f(X)	loss(Y, f(X))
Past performance, trade volume etc. as of Sept 8, 2010	"\$24.50"	"\$24.50"	0
		"\$26.00"	1?
		"\$26.10"	2?

 $loss(Y, f(X)) = (f(X) - Y)^2$ square loss

Performance:

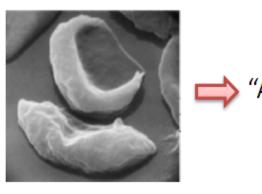
loss(Y, f(X)) - Measure of closeness between true label Y and prediction f(X)

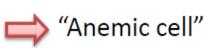
Don't just want label of one test data (cell image), but any cell image $X \in \mathcal{X}$ $(X,Y) \sim P_{XY}$

Given a cell image drawn randomly from the collection of all cell images, how well does the predictor perform on average?

$$\operatorname{Risk} \, R(f) \equiv \mathbb{E}_{XY} \left[\operatorname{loss}(Y, f(X)) \right]$$

Performance: Risk $R(f) \equiv \mathbb{E}_{XY} [loss(Y, f(X))]$





loss(Y, f(X))

$$\mathbf{1}_{\{f(X)\neq Y\}}$$

0/1 loss



$$P(f(X) \neq Y)$$

Probability of Error

$$(f(X) - Y)^2$$

square loss

$$\mathbb{E}[(f(X) - Y)^2]$$

Mean Square Error

Bayes Optimal Rule

Ideal goal: Construct **prediction rule** $f^*: \mathcal{X} \to \mathcal{Y}$

$$f^* = \arg\min_{f} \mathbb{E}_{XY} [loss(Y, f(X))]$$

Bayes optimal rule

Best possible performance:

Bayes Risk
$$R(f^*) \leq R(f)$$
 for all f

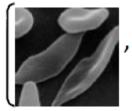
BUT... Optimal rule is not computable - depends on unknown Pxy!

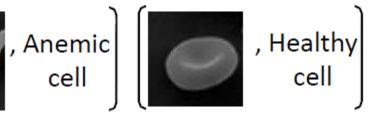
Experience - Training Data

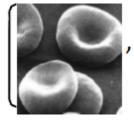
Can't minimize risk since P_{XY} unknown!

Training data (experience) provides a glimpse of P_{xy}

(observed)
$$\{(X_i,Y_i)\}_{i=1}^n \overset{i.i.d.}{\sim} P_{XY}$$
 (unknown) independent, identically distributed









Provided by expert, measuring device, some experiment, ...

Supervised Learning

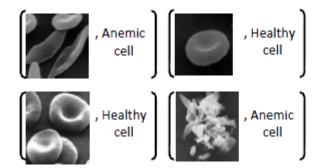
Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

 \equiv Construct **prediction rule** $f: \mathcal{X} \rightarrow \mathcal{Y}$

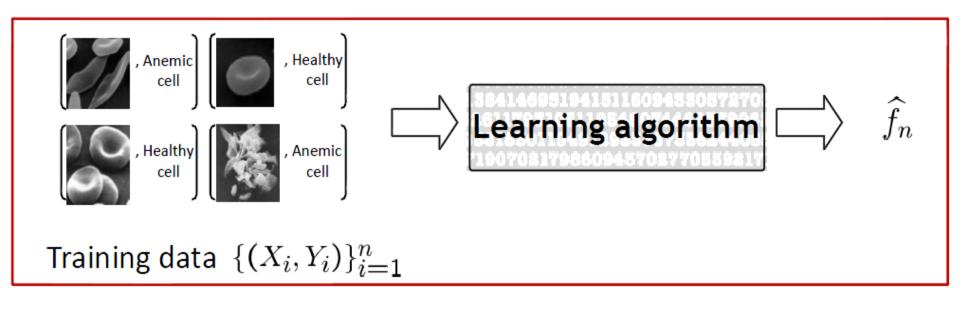
Performance: Risk $R(f) \equiv \mathbb{E}_{XY} [loss(Y, f(X))]$

 $(X,Y) \sim P_{XY}$

Experience: Training data $\{(X_i, Y_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} P_{XY}$ (unknown)



Machine Learning Algorithm

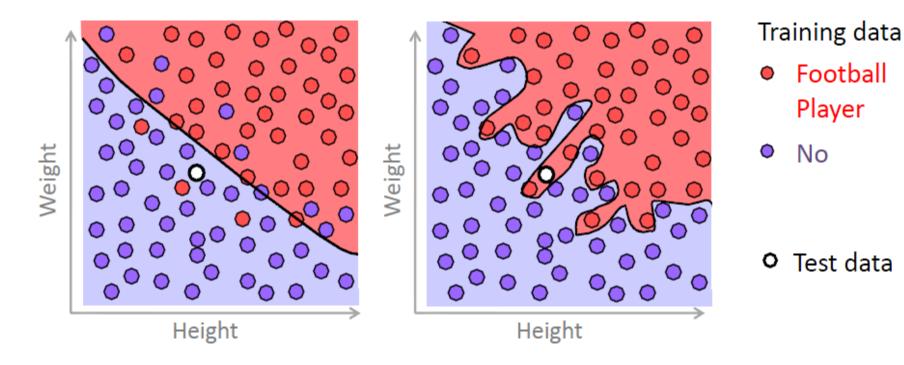


$$\widehat{f}_n$$
 is a mapping from $\mathcal{X} o \mathcal{Y}$ \widehat{f}_n $\left[\begin{array}{c} \widehat{f}_n \end{array} \right]$ = "Anemic cell" Test data X

Note: test data ≠ training data

Issues in ML

- A good machine learning algorithm
 - Does not overfit training data



Generalizes well to test data

More later ...

How to sense Generalization Error?

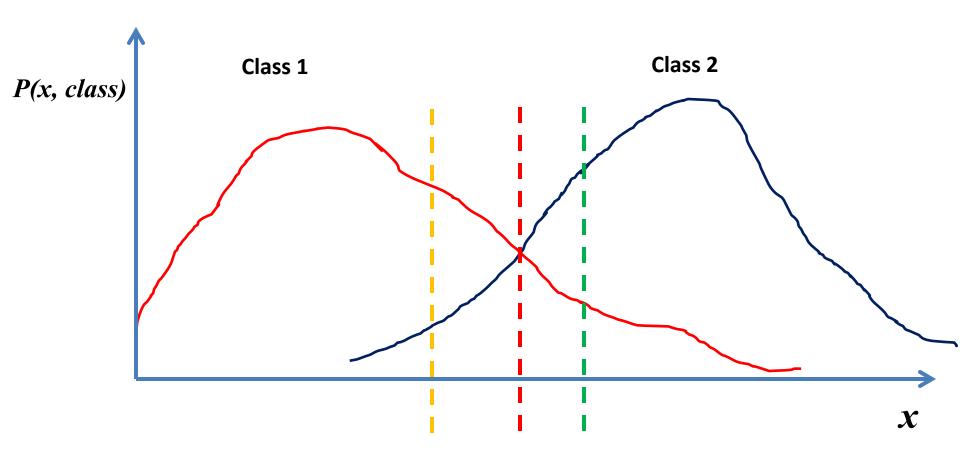
- Can't compute generalization error. How can we get a sense of how well algorithm is performing in practice?
- One approach -
 - Split available data into two sets $\{(X_i, Y_i)\}_{i=1}^n \{(X_i', Y_i')\}_{i=1}^n$
 - Training Data used for training the algorithm

$$\{(X_i,Y_i)\}_{i=1}^n \longrightarrow$$
 Learning algorithm $\longrightarrow \widehat{f}_n$

 Test Data (a.k.a. Validation Data, Hold-out Data) – provides estimate of generalization error

Test Error =
$$\frac{1}{n} \sum_{i=1}^{n} \left[loss(Y_i', \widehat{f}_n(X_i')) \right]$$
 Why not use Training Error?

How to minimize errors?

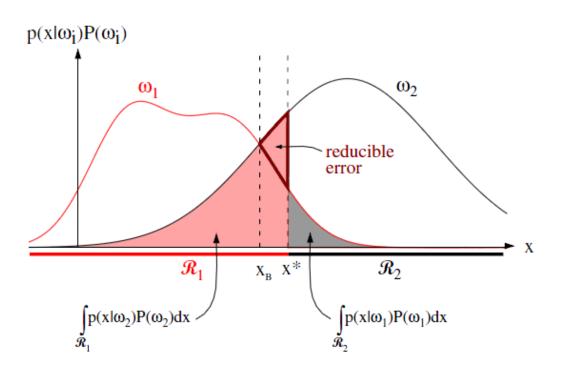


Where to set a threshold on x to make classification in order to minimize classification errors?

Can you get 100% classification accuracy?

Bayes Errors

Calculate the probability of an error – Bayes error



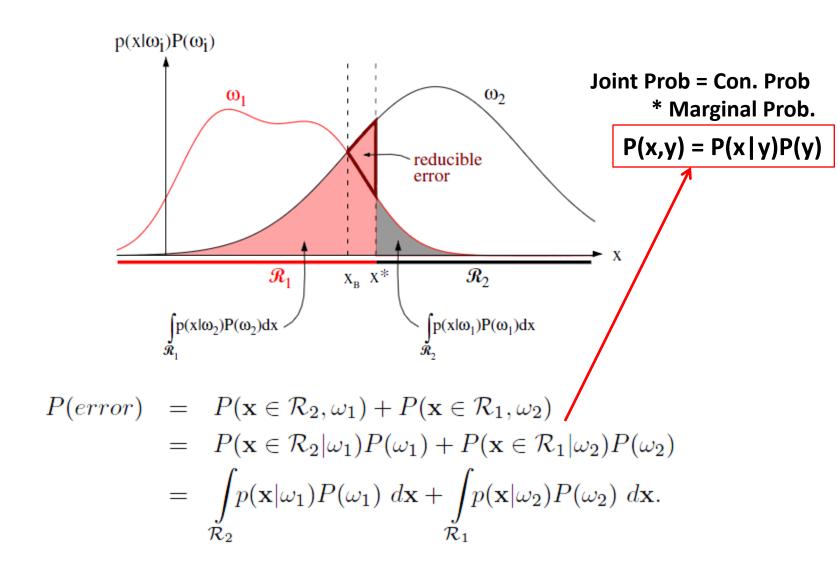
$$P(error) = P(\mathbf{x} \in \mathcal{R}_2, \omega_1) + P(\mathbf{x} \in \mathcal{R}_1, \omega_2)$$

$$= P(\mathbf{x} \in \mathcal{R}_2 | \omega_1) P(\omega_1) + P(\mathbf{x} \in \mathcal{R}_1 | \omega_2) P(\omega_2)$$

$$= \int_{\mathcal{R}_2} p(\mathbf{x} | \omega_1) P(\omega_1) d\mathbf{x} + \int_{\mathcal{R}_1} p(\mathbf{x} | \omega_2) P(\omega_2) d\mathbf{x}.$$

Bayes Errors

Calculate the probability of an error – Bayes error



Bayes optimal classifiers

 Classifier that minimizes the Bayes error is called the Bayes optimal classifier:

• classify
$$x$$
 as
$$\begin{cases} C_0 \text{ if } P(C_0 \mid x) > P(C_1 \mid x) \\ C_1 \text{ oth erwise} \end{cases}$$