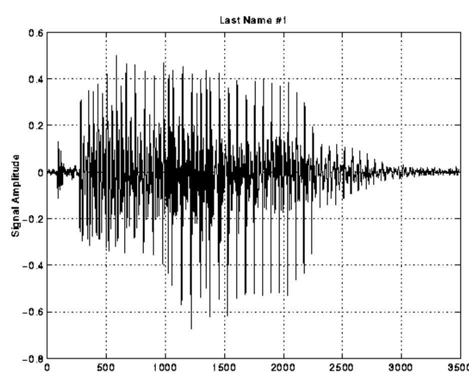
CS425/CS528, Fall 2012
11:10 AM – 12:25 PM
MK 525

Machine Learning

What is Learning? and Why Learn ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)
- But, not always appropriate
  - For example, there is no need to “learn” to calculate payroll

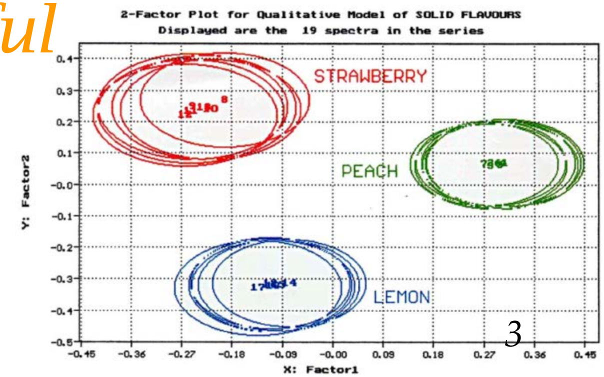


What We Talk About When We Talk About"Learning"

- Learning general models from data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

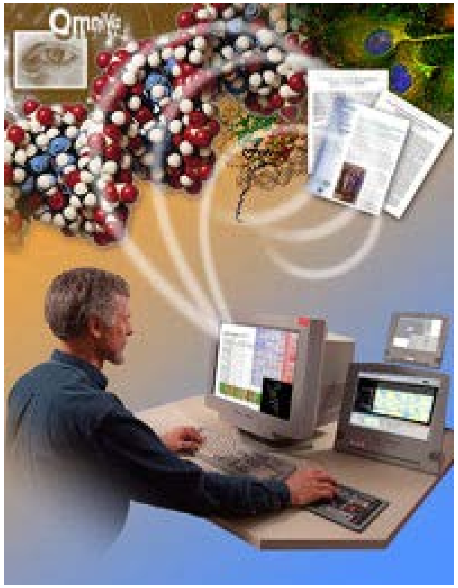
  People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)

- Build a model that is *a good and useful approximation* to the data.



Data Mining: Application of Machine Learning to Large Databases
(also called "Knowledge Discovery in Databases (KDD)")

- Retail: Market basket analysis, Customer relationship management (CRM)
- Finance: Credit scoring, fraud detection
- Manufacturing: Optimization, troubleshooting
- Medicine: Medical diagnosis
- Telecommunications: Quality of service optimization
- Bioinformatics: Motifs, alignment
- Web mining: Search engines
- ...

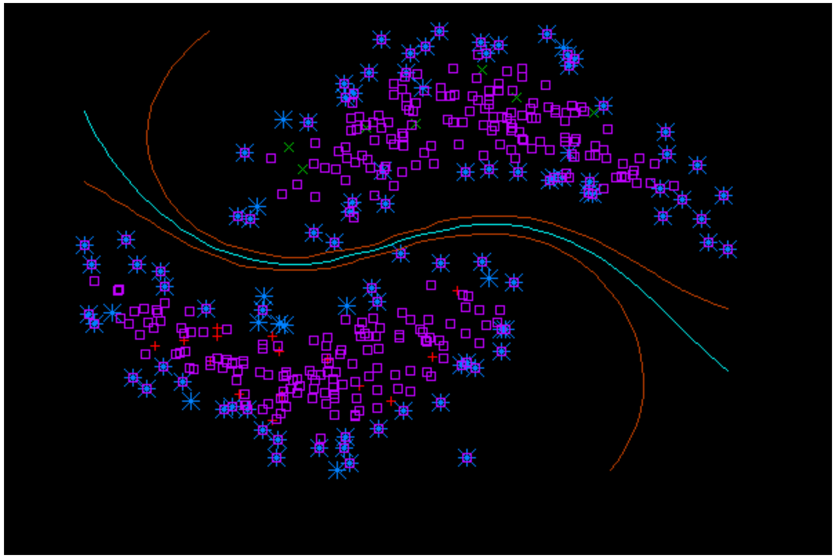


Relevant Disciplines for Machine Learning

- Artificial Intelligence
- Bayesian methods
- Computational complexity theory
- Control theory
- Information theory
- Statistics
- Philosophy
- Psychology
- …

Some Types of Machine Learning

- Learning Associations: Find relationships in the data
- Supervised Learning: We want to learn a mapping from the input to the output; correct values are provided by supervisor
  - Classification
  - Regression
- Unsupervised Learning: We have only input data; we want to find regularities in the data.
- Reinforcement Learning: Learn a policy that maps states to actions.



Learning Associations

- Example: Shopping basket analysis

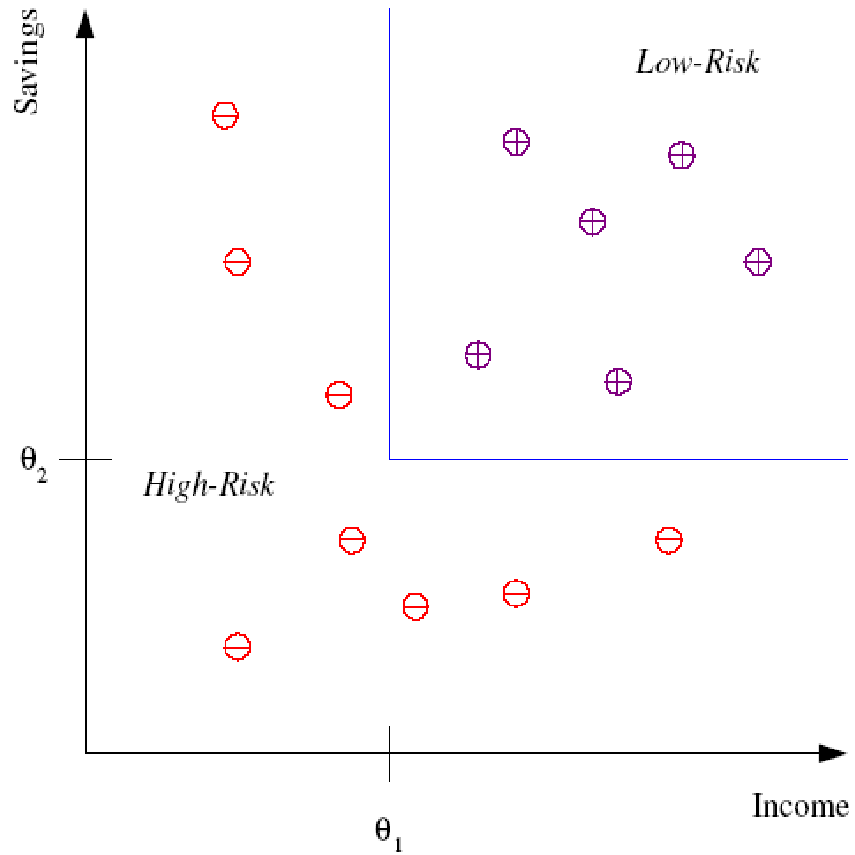
  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

  We learn *Association Rule*: P(chips | soda) =

- Use this Association Rule like this:
  - Target customers who bought X, but not Y
    - Try to convince them to buy Y

Classification (a type of supervised learning)

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2 THEN low-risk ELSE high-risk

- Main application: prediction

Classification: Applications

Also known as: Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to “states” of thought
- Reading text:
- …

Example Pattern Recognition: Face Recognition

Training examples of a person

Test images

AT&T Laboratories, Cambridge UK
http://www.uk.research.att.com/facedatabase.html

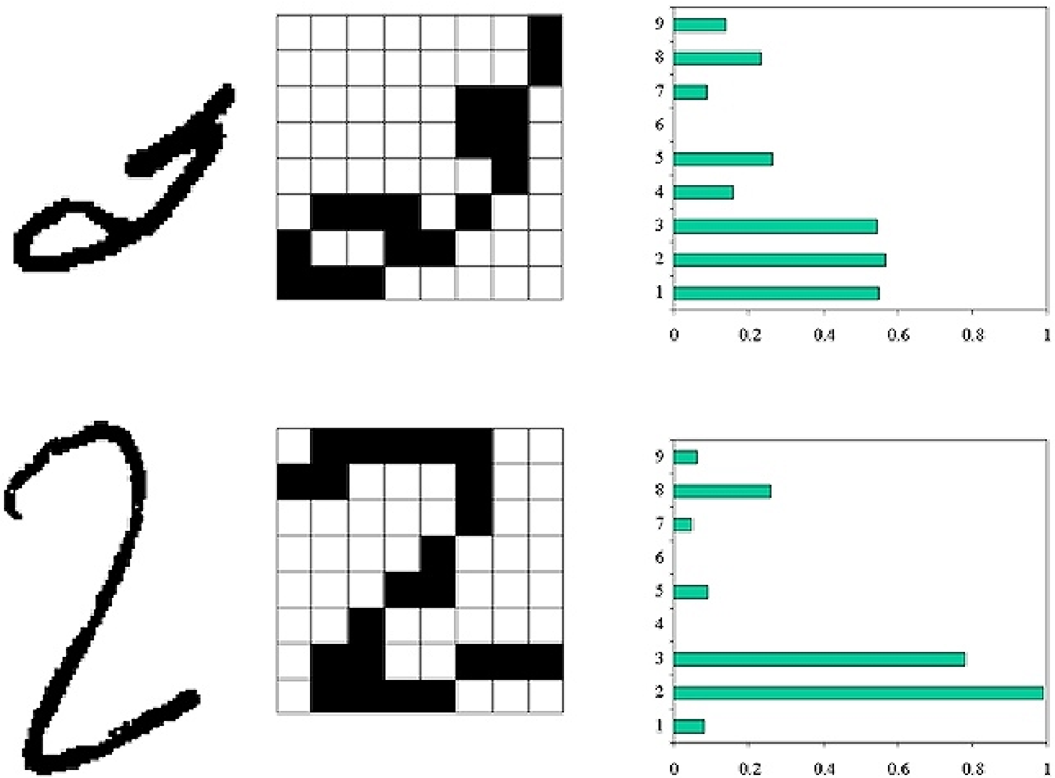
Classification: Applications

Also known as: Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to “states” of thought
- Reading text:
- …

Example Pattern Recognition: Character Recognition

Want to learn how to recognize characters, even if written in different ways by different people

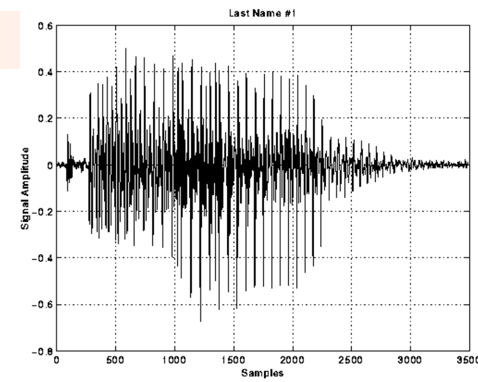


Classification: Applications

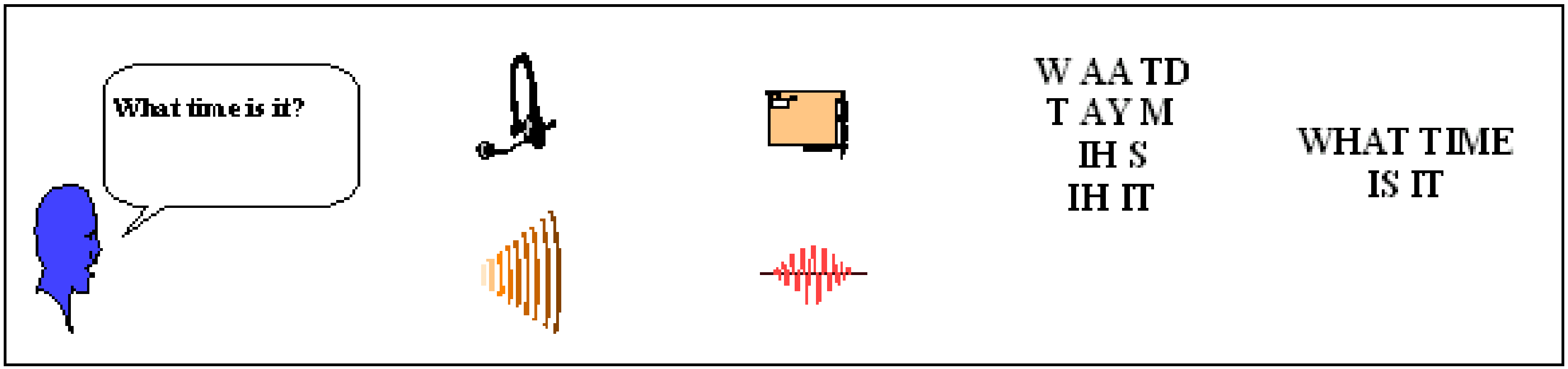
Also known as: Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to “states” of thought
- Reading text:
- …

Example Pattern Recognition: Speech Recognition



USER	MICROPHONE	SOUND CARD	SPEECH RECOGNITION ENGINE	SPEECH-AWARE APPLICATION



User speaks into the microphone.	Microphone captures sound waves and generates electrical impulses.	Sound card converts acoustical signal to digital signal.	Speech recognition engine coverts digital signal to phonemes, then words.	Application processes words as text input.

Classification: Applications

Also known as: Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to “states” of thought
- Reading text:
- …

Example Pattern Recognition: Gesture Recognition

Classification: Applications

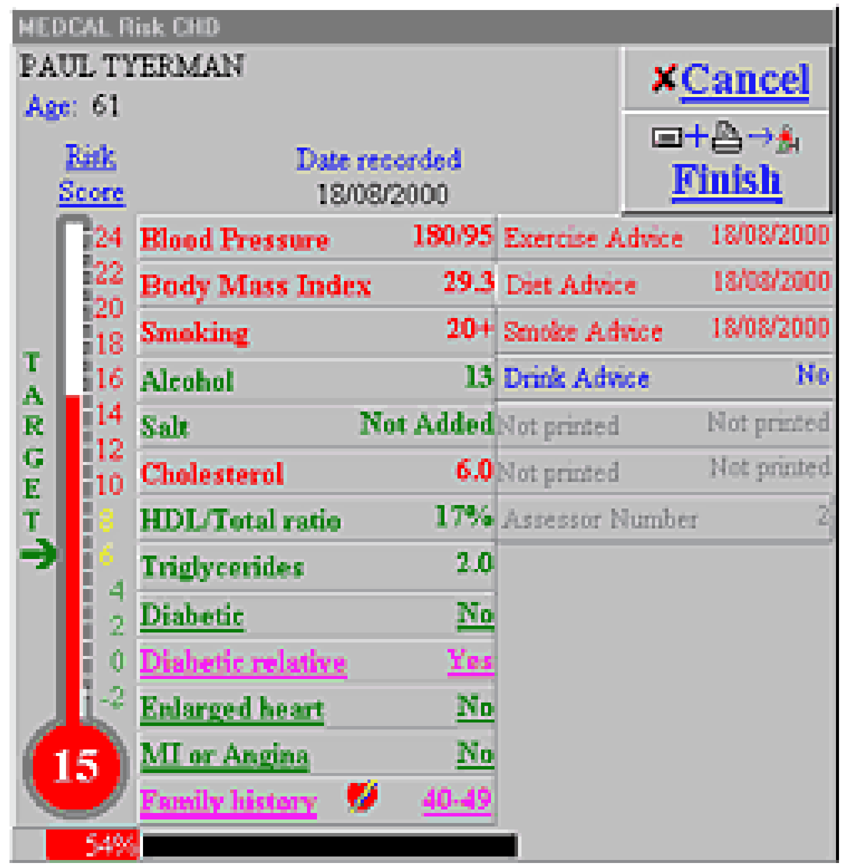
Also known as: Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to "states" of thought
- Reading text:
- …

Example Pattern Recognition: Medical Diagnosis

Inputs: relevant info about patient, symptoms, test results, etc.

Output: Expected illness or risk factors



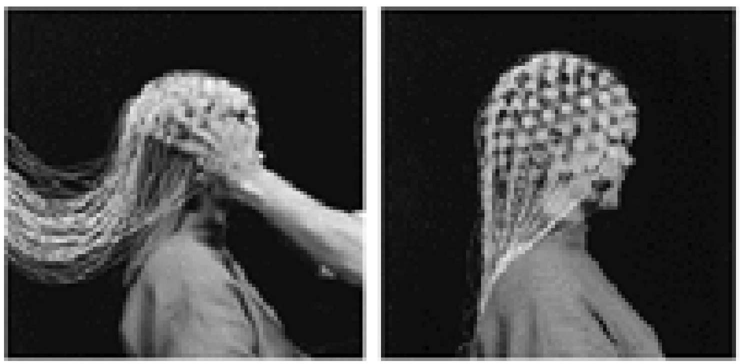
Classification: Applications

Also known as: Pattern recognition

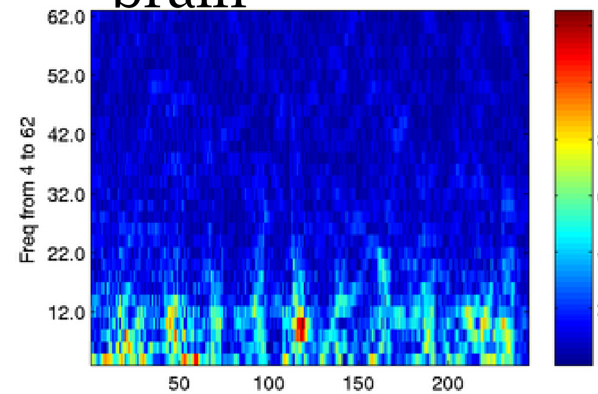
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to “states” of thought
- Reading text:
- …

Example Pattern Recognition: Interpreting Brainwaves

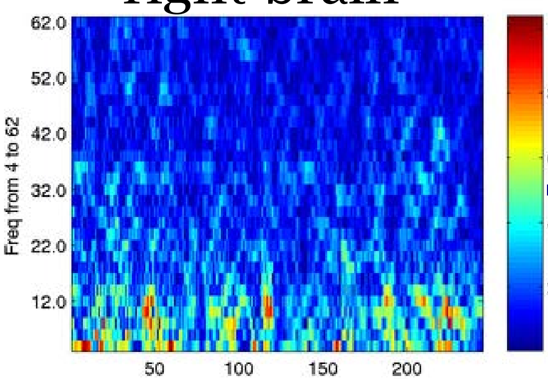
EEG electrodes reading brain waves:



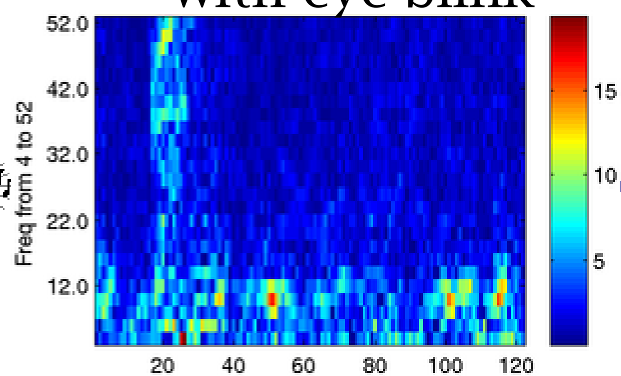
- Rotation task, left brain



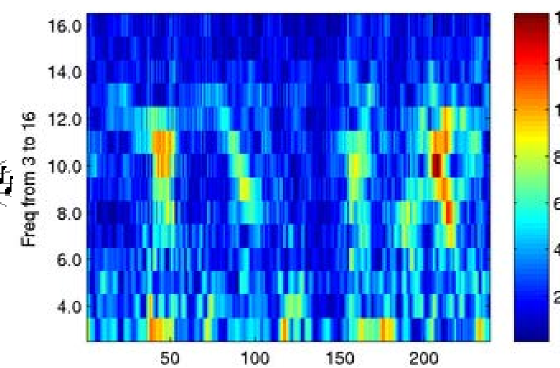
- Rotation task, right brain



- Resting task, with eye blink



- Counting task



Classification: Applications

Also known as: Pattern recognition

- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
  - Use of a dictionary or the syntax of the language.
  - Sensor fusion: Combine multiple modalities; eg, visual (lip image) and acoustic for speech
- Gesture recognition: Different hand shapes.
- Medical diagnosis: From symptoms to illnesses.
- Brainwave understanding: From signals to “states” of thought
- Reading text:
- …

Example Pattern Recognition: Reading text

- Can you read this?
  - Aircndcog to a rseerhcaer at Cbiardmge Urensvitiy, it dsoen't mtetar in waht oderr the letrtes in a wrod are, the olny ipnaotmrt tihng is taht the fsrit and lsat lteter be at the rgiht plcae. The rset can be a toatl mses and you can slitl raed it wutohit porlebm. Tehy spectluae taht tihs is bseuace the hmaun mnid deos not raed erevy leettr by iesltf but the wrod as a whloe. Wtehehr tihs is ture or not is a ponit of deabte.

- Clearly, the brain has learned syntax and semantics of language, including contextual dependencies, to make sense of this ☺

- For fun: Here's a web page where you can create your own jumbled text: http://www.stevesachs.com/jumbler.cgi

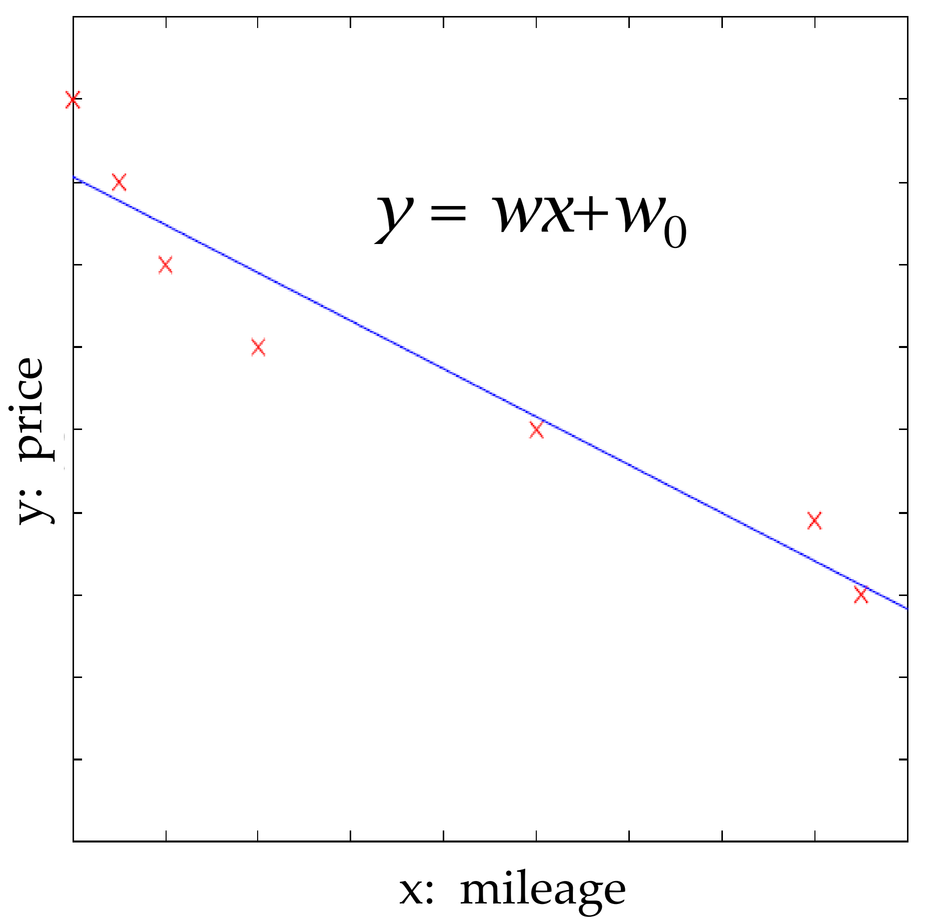
Regression (another type of supervised learning)

- Example:
  - Predict price of a used car
- (Input) x : car attributes (e.g., mileage)

  (Output) y : price
- Our task: learn the mapping from input to output
  - We know basic g ( ) model
  - We want to learn appropriate values for θ parameters that minimize the error in the approximation:

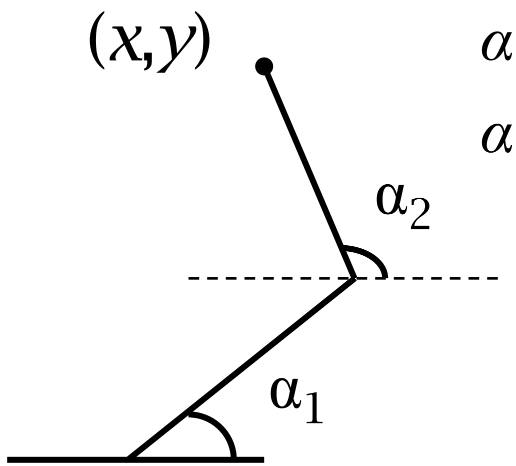
    $y = g(x \mid \theta)$

Here, a linear regression function:



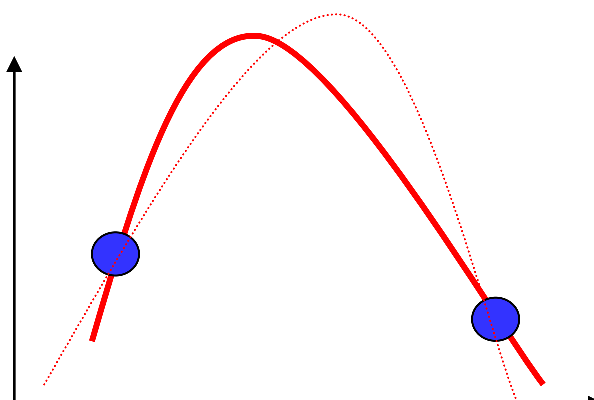
Example Regression Applications

- Navigating a car: Angle of the steering wheel (CMU NavLab)
- Kinematics of a robot arm



$$\alpha_1 = g_1(x,y)$$

$$\alpha_2 = g_2(x,y)$$



- Response surface design
  (using function optimization)

Supervised Learning: Handy Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: We can deduce an explanation about the process underlying the data
- Compression: The rule is simpler than the data it explains
- Outlier detection: We can find instances that do not obey the rule, and are thus exceptions (e.g., to detect fraud)

Unsupervised Learning

- Learning "what normally happens"
- No output available (i.e., we don't know the "right" answer)
- Clustering (density estimation): Grouping similar instances
- Example applications:
  - Customer segmentation in CRM (Customer Relationship Management)
    - Company may have different marketing approaches for different groupings of customers
  - Image compression: Color quantization
    - Instead of using 24 bits to represent 16 million colors, reduce to 6 bits and 64 colors, if the image only uses those 64 colors.
  - Bioinformatics: Learning motifs (i.e., sequences of amino acids that occur repeatedly in proteins)

Reinforcement Learning

- Learning a policy: A sequence of actions to take, given the current state
- No supervised output, but delayed reward is provided
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

Where is Machine Learning Headed?

- Today: tip of the iceberg
  - First-generation algorithms: neural networks, decision trees, regression…
  - Applied to well-formatted databases
  - Budding industry
- Opportunity for tomorrow: enormous impact
  - Learn across full mixed-media data
  - Learn across multiple internal databases, plus the web and newsfeeds
  - Learn by active experimentation
  - Learn decisions rather than predictions
  - Cumulative, lifelong learning
  - Programming languages with learning embedded?

Resources: Journals

- *Journal of Machine Learning Research*
- *Machine Learning*
- *Neural Computation*
- *Neural Networks*
- *IEEE Transactions on Neural Networks*
- *IEEE Transactions on Pattern Analysis and Machine Intelligence*
- *Annals of Statistics*
- *Journal of the American Statistical Association*
- ...

Resources: Conferences

- International Conference on Machine Learning (ICML)
- European Conference on Machine Learning (ECML) and European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)
- Neural Information Processing Systems (NIPS)
- Uncertainty in Artificial Intelligence (UAI)
- Computational Learning Theory (COLT)
- International Joint Conference on Artificial Intelligence (IJCAI)
- International Conference on Neural Networks (Europe)
- ...

Our First Learning Study: Neural Networks

But first, we'll look at some general issues in designing a machine learning system

- For next time, read chapter 1 of Mitchell text