COMP 472: Artificial Intelligence
Machine Learning
Introduction

- Russell & Norvig: Sections 19.1 - 19.2

Next Set of Videos

1. Introduction to ML
2. Naïve Bayes Classification
   a. Application to Spam Filtering
3. Decision Trees
4. (Evaluation
5. Unsupervised Learning)
6. Neural Networks
   a. Perceptrons
   b. Multi Layered Neural Networks

https://medium.com/machine-learning-for-humans/neural-networks-deep-learning-cdad8aeae49b

Today

1. Introduction to ML
2. Naïve Bayes Classification
   a. Application to Spam Filtering
3. Decision Trees
4. (Evaluation
5. Unsupervised Learning)
6. Neural Networks
   a. Perceptrons
   b. Multi Layered Neural Networks

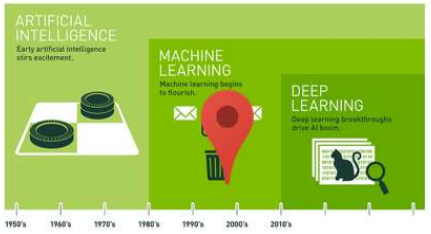
https://medium.com/machine-learning-for-humans/neural-networks-deep-learning-cdad8aeae49b

Remember this slide…

History of AI

- 1980s-2010
- The rise of Machine Learning
  - More powerful CPUs-> usable implementation of neural networks
  - Big data -> Huge data sets are available
    - document repositories for NLP (e.g. emails)
    - billions on images for image retrieval
    - billions of genomic sequences, …

😃 Rules are now learned automatically !



2011: Watson wins at Jeapardy!

40

Motivation

- Too many to list here!
  - Recommender systems (eg. Netflix)
  - Pattern Recognition (eg. Handwriting recognition)
  - Detecting credit card fraud
  - Computer vision (eg. Object recognition)
  - Discovering Genetic Causes of Diseases
  - Natural Language Processing (eg. Spam filtering)
  - Speech Recognition / Synthesis
  - Medical Diagnostics
  - Information Retrieval (eg. Image search)
  - Learning heuristics for game playing
  - …
  - *Oh… I'm out of space*

What is Machine Learning?

- Learning = crucial characteristic of an intelligent agent 重要的，决定性的

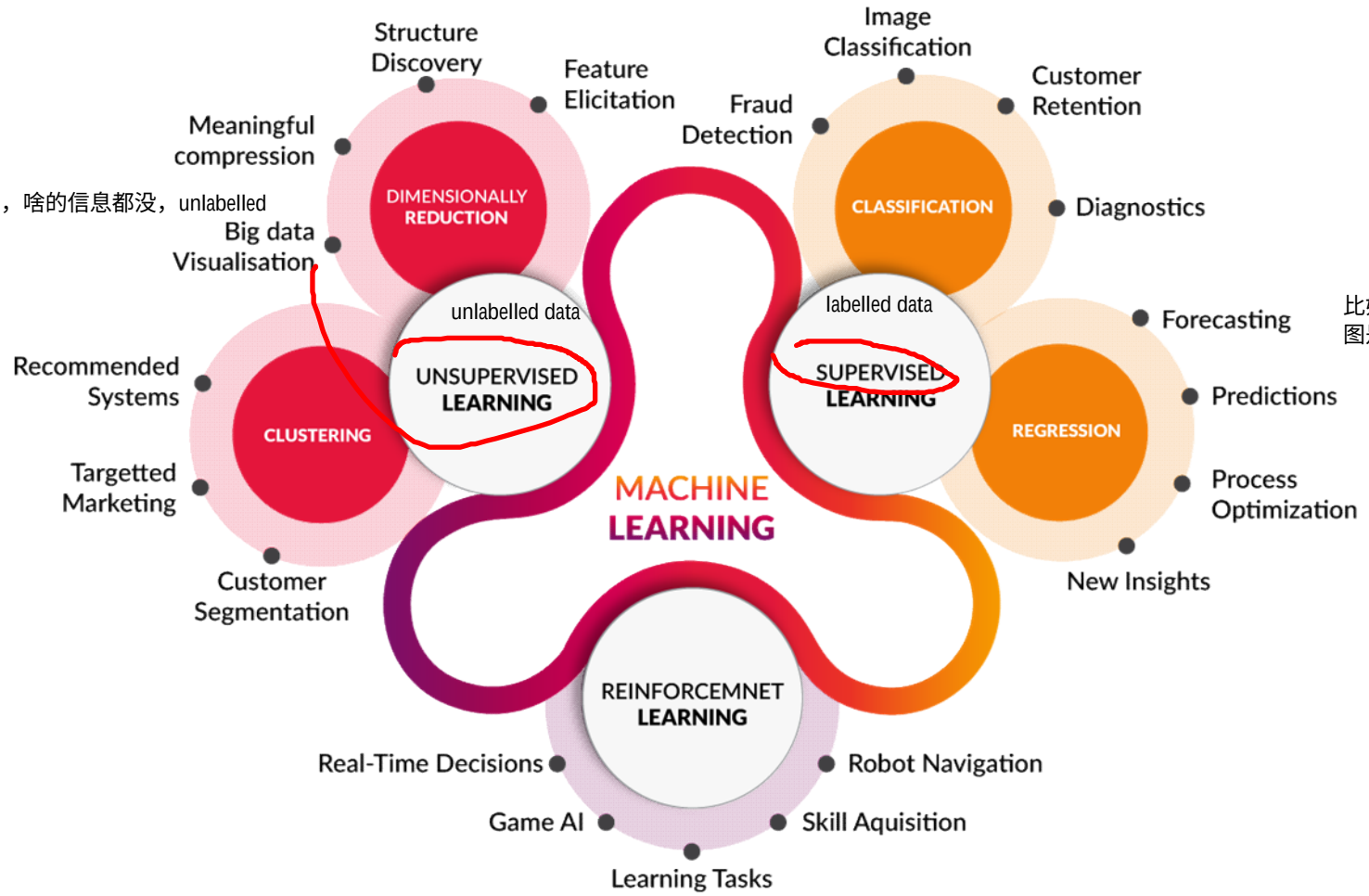
- ML
  - Constructs algorithms that learn from data
  - ie perform tasks that were not explicitly programmed and improve their performance the more tasks they accomplish 完成越多task越有效
    - generalize from given experiences and are able to make judgments in new situations

概括

不是learning by heart ,而是genealize

Types of Machine Learning

只有图片，啥的信息都没，unlabelled

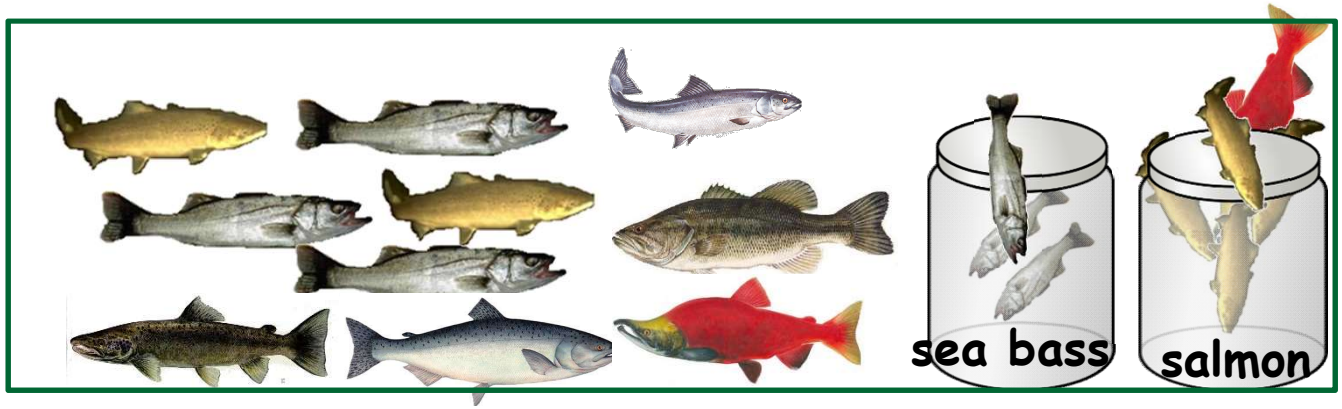


比如一组图，分好类了，哪张图是猫，哪张图是狗，这种就是labelled data

http://www.cognub.com/index.php/cognitive-platform/

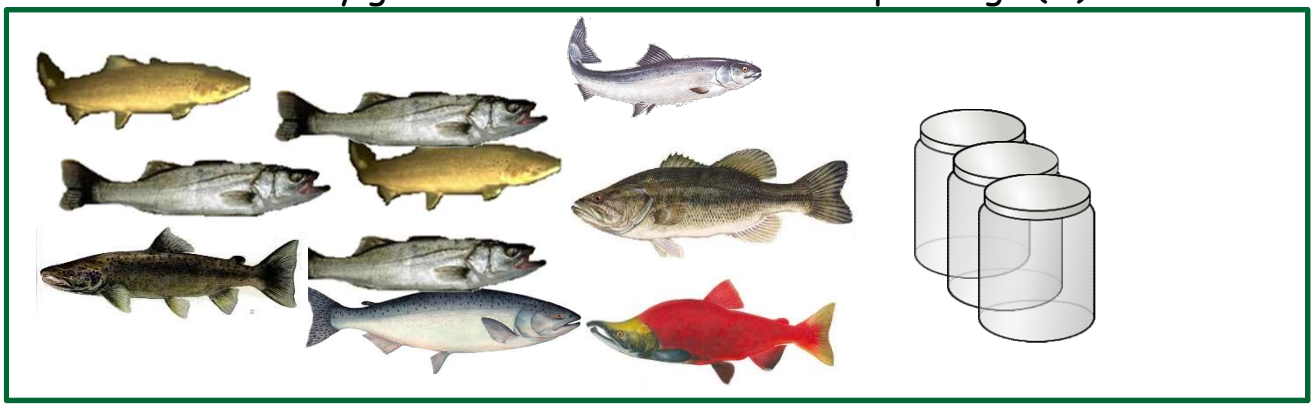
Types of Machine Learning

- Supervised learning
  - We are given a training set of (X, f(X)) pairs
  - X = <color, length>



?

- Unsupervised learning 主要区别在于f(x)
  - We are only given the Xs - not the corresponding f(X)



?

Types of Learning

- In **Supervised** learning
  - We are given a training set of (X, f(X)) pairs

actual label

big nose	big teeth	big eyes	no moustache	f(X) = not person
small nose	small teeth	small eyes	no moustache	f(X) = person
small nose	big teeth	small eyes	moustache	f(X) = ?

- In **Reinforcement** learning
  - We are not given the (X, f(X)) pairs

一开始不给f(x)，但是如果f(x)对了，给你reward

small nose	big teeth	small eyes	moustache	f(X) = ?

  - But we get a reward when our learned f(X) is right, and we try to maximize the reward
  - Goal: maximize the nb of right answers

- In **Unsupervised** learning
  - We are only given the Xs - not the corresponding f(X)

not given actual label

big nose	big teeth	big eyes	no moustache	*not given*
small nose	small teeth	small eyes	no moustache	*not given*
small nose	big teeth	small eyes	moustache	f(X) = ?

  - No *teacher* involved / Goal: find regularities among the Xs (clustering)
  - Data mining

Logical Inference 逻辑推理

- Inference: process of deriving new facts from a set of premises

  前提

- Types of logical inference:
  1. Deduction 顺推
  2. Abduction 逆推
  3. Induction

Deduction

- aka Natural Deduction 自然推理
- Conclusion follows necessary from the premises. 结论需要前置条件
- From A $\Rightarrow$ B and A, we conclude that B
- We conclude from the general case to a specific example of the general case
- Ex:

All men are mortal.
Marcus is a man.

Marcus is mortal.

Abduction

逆推 假设

- Conclusion is one hypothetical (most probable) explanation for the premises
- From $A \Rightarrow B$ and B, we conclude A
- Ex:

  Drunk people do not walk straight.
  John does not walk straight.

  John is drunk.

- Not sound… but may be most likely explanation for B
- Used in medicine…
  - in reality… disease $\Rightarrow$ symptoms
  - patient complains about some symptoms… doctor concludes a disease

Induction

归纳

- Conclusion about all members of a class from the examination of only a few member of the class.
- From $A \wedge C \Rightarrow B$ and $A \wedge D \Rightarrow B$, we conclude $A \Rightarrow B$
- We construct a general explanation based on a specific case.
- Ex:

  All CS students in COMP 472 are smart.

  All CS students on vacation are smart.

  All CS students are smart.

- Not sound
- But, can be seen as hypothesis construction or generalisation

Inductive Learning

- = learning from examples
- Most work in ML
- Training set is given:
  - that includes examples of already classified examples
  - i.e. pairs of (X, f(X))
  - Ex:

supervised learning,
f(x)给出来了

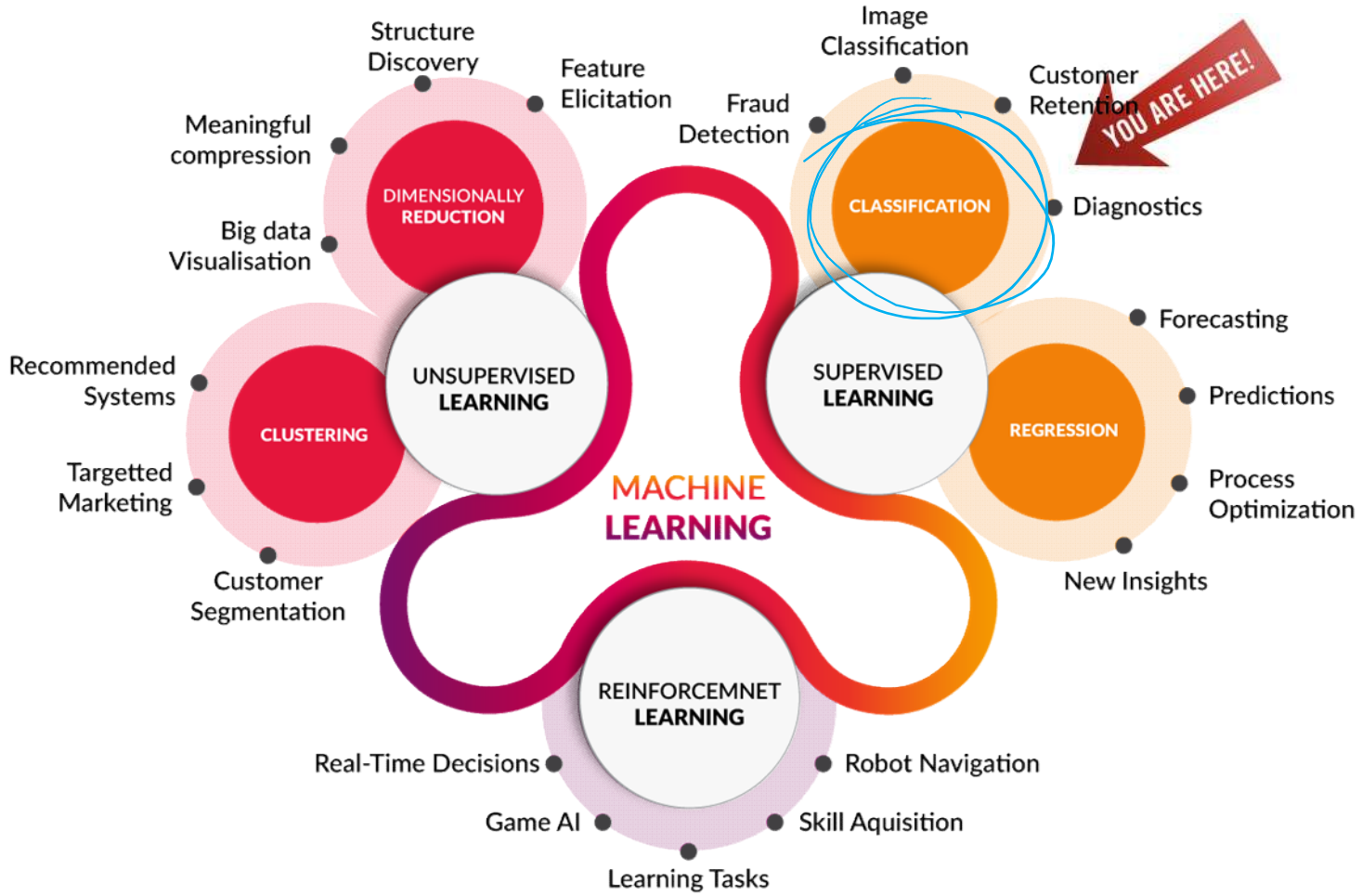
small nose	big teeth	small eyes	moustache	tiger

  - used to "train" a model
- Test set:
  - that includes examples of non classified examples
  - for which the model must to make accurate predictions

big nose	big teeth	small eyes	moustache	?

- Can be seen as learning a function
  - the model uses the training set to find an estimate of the function f(X) given X
  - so that given a new instance X it has never seen before, the model can predict f(X)

Types of Machine Learning

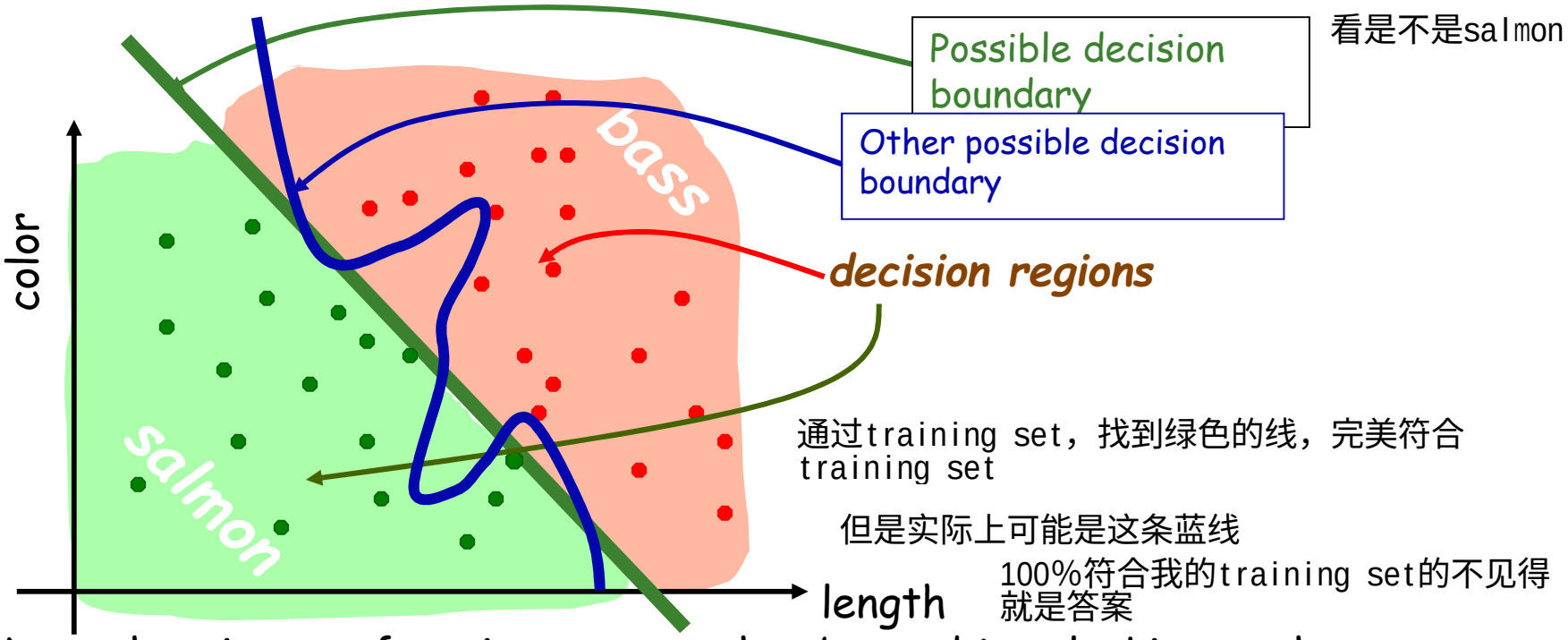


http://www.cognub.com/index.php/cognitive-platform/

Example

- Given pairs (X,f(X)) (the training set – the data points) training set 就是这些点
- Find a function that fits the training set well

fish(color：xx,length:xx) color是x length是y，



- Note: choosing one function over another beyond just looking at the training set is called **inductive bias** (eg. prefer "smoother" functions)

指的是更愿意选择平滑曲线

Inductive Learning Framework

- Input data are represented by a vector of features, X
- Each vector X is a list of (attribute, value) pairs.
  - Ex: `X = [nose:big, teeth:big, eyes:big, moustache:no]`
- The number of attributes is fixed (positive, finite)
- Values can be
  1. categorical (aka nominal) - unordered values (范畴的)
     - eg. "convertible", "minivan", "SUV", ... "yes", "no" (分类型的，例如说小面包，suv这种)
  2. ordinal - ordered values (顺序的)
     - e.g., "large", "medium" or "small" (有顺序的value)
  3. numerical – ordered (数字的)
     - e.g. frequency of a word in an email (40, 55)
     - e.g. height of a person (1.7, 1.6)

Note: *attribute == feature*

- Each example can be interpreted as a point in a n-dimensional feature space, where n is the number of attributes

Example

x						f(x)
has-hair?	has-scales?	has-feathers?	flies?	lives in water?	lays eggs?	
1	0	0	0	0	0	Dog
1	0	0	0	0	0	Cat
1	0	0	1	0	0	Bat
1	0	0	0	1	0	Whale
0	0	1	1	0	1	Canary
0	0	1	1	0	1	Robin
0	0	1	1	0	1	Ostrich
0	1	0	0	0	1	Snake
0	1	0	0	0	1	Lizard
0	1	0	0	1	1	Alligator

Real ML applications typically require hundreds, thousands or millions of examples

source: Alison Cawsey: The Essence of AI (1997).

Techniques in ML

概率统计的方法

- Probabilistic Methods
  - output a probability associated with the result of the classification 分类后输出一个概率
- Decision Trees 使用可以识别的特征
  - Use only discriminating features as questions in a big if-then-else tree
- Neural networks
  - Also called parallel distributed processing or connectionist systems
  - Intelligence arise from having a large number of simple computational units
- …

比如说前面18页，通过所有的属性，来得到结果

Today

1. Introduction to ML 
2. Naïve Bayes Classification
   a. Application to Spam Filtering
3. Decision Trees
4. (Evaluation
5. Unsupervised Learning)
6. Neural Networks
   a. Perceptrons
   b. Multi Layered Neural Networks

Up Next

1. Introduction to ML
2. Naïve Bayes Classification
   a. Application to Spam Filtering
3. Decision Trees
4. (Evaluation
5. Unsupervised Learning)
6. Neural Networks
   a. Perceptrons
   b. Multi Layered Neural Networks