Supervised Learning

Dr. Jianlin Cheng

University of Missouri, Columbia

Slides Adapted from Book and CMU, Stanford Machine Learning Courses Fall, 2011

Syllabus

- Course web site: <u>http://people.cs.missouri.edu/~chengji/supervised_lea</u> <u>rning/</u>
- Location: EBW 355; Time: TuTh 4:00 pm 5:15 pm; Office Hours: TuTh 3:00 pm - 4:00 pm
- **Text Book**: Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2007
- Grading: assignment (40%), group project report (30%), group project representation (30%); grade scale (A+, A, A-, B+, B, B-, C+, C, C-, and F)
- Questions / Assignment submission: <u>mumachinelearning@gmail.com</u>

Topics

- Introduction and Bayes optimal learning rule
- MLE and MAP (parametric learning)
- Generative VS discriminative methods
- Nonparametric methods
- Model selection
- Boosting and bagging
- Support vector machines
- Graphical models (emphasized)
- Semi-supervised 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>



Decoding thoughts from brain scans





Rob a bank ...

Home » Health & Wellness

Brain Scans

Brain Scans: Are You a Criminal?



More:

Published February 07, 2007 by: Andrea Okrentowich View Profile | Follow | Add to Favorites

Brain Scan Disposition Defendant Criminal Behavior

MRI Scans as Courtroom Evidence



Stock Market Prediction



Document classification



Sports Science News

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.

Spam/ Not spam

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

===	Natural	WeightL0SS	Solution	===
-----	---------	------------	----------	-----

Vital Acai is a natural WeightL0SS 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.



- 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 ...





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

Robot assistant?

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



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!



Object Recognition

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







Face Recognition









Face Recognition









```
Latttaa
itate a transmission of the state the available transmission of the state as the state as a transmission of the state as a trank of the state as a transmission of the state as a transm
: ttttttttagggcccaataaaagcgcttttgtggcggcttgatagattatcacttggtttcggtggctagccaagtggctttcttctgtccgacgcacttaattgaattaaccaaacaacgagcgtggccaattcgtattatcgctg
                                                                                                       <u>mtqqqqqqtcaqtcaqtetaqqqaqaqqqqqcqcccaccaqtcqatcacqaaaacqqcqaattecaaqcqaaacqqaaacqqaqcqaqcqaqcqaqc</u>
gteettegageegattataeaegaeageaaaatattteagtegegettgggeaaaaggettaageaegaeteeeagteeeeettaeatttgtetteetaageeeetggageeaetateaaaettgttetaegettgeaetgaaaata
:caaagtaaacaatcaaaaagaccaaaaaacaataaccagcaccagagtcgaacatcagtgaggcattgcaaaaatttcaaagtcaagtttgcgtcgtcatcgcgtctgagtccgatcaagccgggcttgtaattgaagttgttgatgi
:actggattgtggggaattetggteageataettaacageageeegetaattaageaaaataaacatateeagaatgegaeggegeeateateetgtttggggaatteaatteggggeagategtttaatteaatteaaaggt
g cattetaatgaagaettggagaagaettaegttatatteagaeeategtgegatagaggatgagteattteeatatggeegaaatttattatgtttaetategttttttagaggtgtttttttggaettaeeaaaagaggeatttgttt
gtacttgaggaattggegategtagegattteeeeegeegtaaagtteetgateetegttgtttttgtaeateataaagteeggattetgetegeegaagatggggaaegaagetgeegaagetgeegaagetgeetgaggtgetgg
:cccagctggataaccttgctgtacagatcggcatctgcctggagggcacgatcgaaatccttccagtggacgaacttcacctgctcgctgggaatagcgttgttgtcaagcagctcaaggagcgtattcgagttgacgggctgcacca
tgegteetatgegaggaatggtetttaggttetttatggeaaagttetegeetegettgeeeageeggtaegttettggtgatetttaggaafaateetggaetaetftegtetgeetggettatggeeaeaagaeeeaeeaagag
; as a actga acta a cot gaceta a type cata atta a gtg cata cata cata catta catta catt type a tata cata construction of the second sec
cagcatattttacgtaaaaacaaaacggtaatgegaacataacttatttattggggcccggaccgcaaaccggccaaacgegtttgcaccgataaaaacataagggcaaqaaaaaattgttaagctgttgtttatttttgcaatcga
: taaccaaaaaacagcagataaacaaaaacggcaaagaaaattgccacagagttgtcacgctttgttgcacaaacatttgtgcagaaaagcttttagccattattagccattattcagctcgctggcagcacttgcgaatgi
geggacteccetegaacgetetegaacteagtggettteeageeggeeegetgggeegetegeeeaetggaeeggtatteeegggeeaetgtaeegeaeegetataateetegeeagaeteggegetgataaggeeeaatg
t cege aggegt et at that gee a agga cegt tet tet te aget t tege et gas gas a tege agget a tege aggegt a tege aggegt at tege aggegt act gas a tege aggegt at tege aggegt aggegt at tege aggegt at tege aggegt at tege aggegt at tege aggegt aggegt aggegt at tege aggegt agggt agggt agggt agggt agggt agggt aggegt agggt agggt agggt a
tteatgetgacatteatatteatteattegttttttgtttttgacgacggactgaaaagtgeggagagaaaceeaaaacagafgegegeaaagegeegttaatatgegaaceeatettgaagttateacaacaeeatateea
itatecatateaatateaatategetattattategetetgetgeteaagtatteagegetgegetagattegaefgattgaategageteaatagaeteaatageteeaetegaeagtgetgetgetgetgetgetgetgetgetgattgaategaegeteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaatageteaa
gagtaaacgaggegtatgegcaacetgeacetggeggaegeggegtatgegeaatgtgeaattegettaeettetegtf_{g}egggteaggaaeteeeagatgggaatggeegatgatetgaatgtggaaggegeeeageag
gattactttegeegeagtegtegtegttgetgettettatgttgegtaeteegeaetaeaeggagagtteaggggattegtgeteegtgatetgtgateegtgtte<math>gtgggteaattgegttegttgtgttgtgtaaeettegt
gteettegageegattataeaegaeageaaaatattteagtegegettgggeaaaaggettaageaegaeteeeagteeeeettaeatttgtetteetaageeeetggageeaetateaaaettgttetaegettgeaetgaaaata
:caaagtaaacaatcaaaaagaccaaaaaacaataaccagcaccagagtcgaacatcagtgaggcattgcaaaaatttcaaagtcaagtttgcgtcgtcatcgcgtctgagtccgatcaagccgggcttgtaattgaagttgttgatgi
:actggattgtggggaattetggteageataettaacageageeegetaattaageaaaataaacatateeagaatgegaeggegeeateateetgtttggggaatteaatteggggeagategtttaatteaatteaaaggt
: tge agtgg agtgg acctgg tea aaga age accept to a tge tee atteg attee a tee gg a tge gg a tge gata aaga agg et et gate ca age a a tge gattte gattte catttgg tt tg tattta eg t a tge gat a tge gat agg agt gg a tge gat gat gg a tge gat gg a tge gat gat gg a tge gat gg a tge gat gat gat gg a tge gat gg a tge gat gat gg a tge gat gg
geattetaatgaagaettggagaagaettaegttatatteagaeeategtgegatagaggatgagteattteeatatggeegaaatttattatgtttaetategttttttagaggtgttttttggaettaeeaaaagaggeatttgttt
gtacttgaggaattggcgatcgtagcgattteeceegeegtaaagtteetgateetegttgtttttgtacateataagteeggattetgetegtegeegaagatggggaacgaagetgeegaagetgeegaagetgeetgaggtgetgg
:cccagetggataaccttgetgtacagateggcatetgectggagggcacgategaaatectteccagtggacgaactteacctgetegggaatagegttgttgtcaagcageteaaggagegtattegagttgacgggetgeacca
```

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, NMMs, graphical models, active learning, reinforcement learning...
- Covers algorithms, theory and applications
- It's going to be fun and hard work I

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



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

Supervised Learning - Classification

Feature Space \mathcal{X}

Words in a document

Label Space ${\mathcal Y}$

...

"Sports" "News" "Science"





"Anemic cell" "Healthy cell"

Discrete Labels

Supervised Learning - Regression

Feature Space $\mathcal X$

Label Space ${\mathcal Y}$



Supervised Learning problems



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
 - ...

- ML apps. All software apps.
- 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 $f: X \rightarrow Y$
- Set of function hypotheses $H=\{h \mid h: X \rightarrow Y\}$

Given:

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

Determine:

• Hypothesis $h \in H$ 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 $f : \mathcal{X} \to \mathcal{Y}$





"Anemic cell (0)"







Performance:



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

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.	"\$24.50"	"\$24.50"	0
as of Sept 8, 2010		"\$26.00"	1?
		"\$26.10"	2?
	oss(Y, f(X)) = (f($(X) - Y)^2$ sq	uare loss

Performance:

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?

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

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



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)
$$\{(X_i, Y_i)\}_{i=1}^n \stackrel{i.i.d.}{\sim} P_{XY}$$
 (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** $f : \mathcal{X} \to \mathcal{Y}$

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

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



Machine Learning Algorithm



 \widehat{f}_n is a mapping from $\mathcal{X} \to \mathcal{Y}$



= "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 $\longrightarrow \hat{f}_n$

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

Test Error =
$$\frac{1}{n} \sum_{i=1}^{n} \left[loss(Y'_i, \hat{f}_n(X'_i)) \right]$$
 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?

Bayes Error

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 Error

Calculate the probability of an error – Bayes error



Classifiers and Bayes error

- A classifier h is a mapping from feature vectors x to class labels {C₀, C₁}
- The Bayes error of h is the probability of a misclassification:

$$\int_{x \in H_0} P(C_1 \mid x) p(x) \, dx + \int_{x \in H_1} P(C_0 \mid x) p(x) \, dx$$

$$x \in H_0$$
Area that h
classifies x as C_0

Bayes optimal classifiers

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

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