

# Introduction to Machine Learning

Inas A. Yassine, PhD

Assoc. Prof., Systems and Biomedical Engineering Department, Cairo University

inas.yassine@eng.cu.edu.eg

# Outlines



- What is machine learning
  - By example
  - Definition
- Medical Informatics
  - Definition
  - Applications
    - Clinical decision making
    - Content based Medical Image Retrieval

#### **Introduction to Machine Learning**



# What is Machine Learning

(By Example)

**Introduction to Machine Learning** 



# Classification

From Data to Discrete Classes

**Introduction to Machine Learning** 

# Spam Filtering

#### data

#### 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 +

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: <u>10615-announce@cs.cmu.edu</u>. You can contact the instructors by emailing: <u>10615-instructors@cs.cmu.edu</u>

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

Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang show details 9:52 PM (1 hour ago) 5 Reply

=== 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 WeightLOSS \* Increased metabolism - BurnFat & calories easily! \* Better Mood and Attitude \* More Self Confidence \* Cleanse and Detoxify Your Body \* Much More Energy \* BetterSex.Life A Natural Colon Cleanse



#### Prediction

#### Spam

#### VS.

#### Not Spam

#### **Introduction to Machine Learning**

### Face Recognition





Example training images for each orientation



#### **Introduction to Machine Learning**

### Weather Prediction









**Introduction to Machine Learning** 



# Regression

Predicting a numeric Value

**Introduction to Machine Learning** 

### **Stock Market**





**Introduction to Machine Learning** 



### Weather Prediction ... Revisited



#### **Introduction to Machine Learning**



# Ranking

**Comparing Items** 

**Introduction to Machine Learning** 



### Web Search

| Coorle            | learning to much  |
|-------------------|---|
| Google            |   |
|                   | learning to rank  |
|                   | learning to rank for information retrieval I'm Feeling Lucky »                        |
| Search            | learning to rank using gradient descent   |
|                   | learning to rank <b>tutorial</b>  |
| Web               | Learning to rank - Wikipedia, the free encyclopedia                                   |
| Imagaa            | en.wikipedia.org/wiki/ <b>Learning_to_rank</b>  |
| images            | Learning to rank or machine-learned ranking (MLR) is a type of supervised or          |
| Maps              | semi-supervised machine learning problem in which the goal is to automatically        |
| Videos            | Applications Feature vectors Evaluation measures Approaches                           |
|                   | Yahoo! Learning to Rank Challenge   |
| News              | learningtorankchallenge.yahoo.com/  |
| Shopping          | Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce |
|                   | determination were some of the highlights of the first ever Yahoo!                    |
| More              |   |
|                   | [PDF] Large Scale Learning to Rank  |
| Manhattan, NY     | www.eecs.tutts.edu/~dsculley/papers/large-scale-rank.pdf                              |
| 10012             | by D Sculley - Cited by 24 - Related articles   |
| Change location   | 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               |
| Show search tools |   |
| Show search tools | Microsoft Learning to Rank Datasets - Microsoft Research                              |
|                   | 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                        |
|                   | 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             |

#### **Introduction to Machine Learning**



# Given Image, Find Similar Images



. Search mode: Theme 2. Find similar by Color / Texture



..... OR ..... 2. Find similar by Color / Texture



----- OR -----2. Find similar by Color / Texture



2. Search mode: Color / Texture



---- OR -----2. Find similar by Color / Texture



flickr

1. Find similar by Theme ..... OR .... 2. Find similar by Color / Texture



----- OR ------2. Find similar by Color / Texture

1. Find similar by Theme

2. Find similar by Color / Texture

----- OR ------



----- OR -----2. Find similar by Color / Texture

1. Find similar by Theme

2. Find similar by Color / Texture

----- OR ------



----- OR ------2. Find similar by Color / Texture



1. Find similar by Theme ----- OR -2. Find similar by Color / Texture





----- OR ------2. Find similar by Color / Texture





····· OR ·····









# **Collaborative Filtering**

**Introduction to Machine Learning** 



### **Recommendation Systems**

| amazon   | David's Ama   | zon.com   Today's D   | Peals   Gift Cards   Sell   Help  | X                                 | Back-1         | Shop now       | s<br>Savings |
|--|---------------|---|---|-----------------------------------|----------------|----------------|--------------|
| shop by Solution Solu | Search        | ooks 🔻  | c   | Go Hello, David<br>Your Account - | Try<br>Prime - | ↓<br>Cart -    | Wish<br>List |
| our Amazon.com Your E  | Browsing Hist | ory Recommended F   | For You Amazon Betterizer Improve Your Recommendations  | Your Profile Learn More           |                |                |              |
| r Amazon.com > Reco  | ommende       | d for You > Books   | Subjects > Science & Math > History & Philos  | ophy                              |                |                |              |
| t For Today  | Th            | ese recommenda  | tions are based on <u>items you own</u> and more.   |                                   |                |                |              |
| owse Recommended   | vie           | w: All   <u>New Relea</u>   | ases   <u>Coming Soon</u>   |                                   |                |                |              |
| commendations<br>story & Philosophy<br>listory of Science<br>hilosophy of Biology<br>hilosophy of Medicine   | 1.            | CAUSALITY<br>CAUSALITY<br>TODEA PEARL                               | Causality: Models, Reasoning and Inference<br>by Judea Pearl (September 14, 2009)<br>Average Customer Review: ★★★★★★★ ⓒ (10)<br>In Stock<br>List Price: \$50.00<br>Price: \$32.49<br>61 used & new from \$28.00<br>ot Interested ⓒ ★★★★★★★★ Rate this Item<br>ause you purchased Probabilistic Graphical Models and more (1         | Add to Cart Add to Wisi           | List           |                |              |
|  | 2.            | LOCK INSDEL<br>THE'L ADY<br>AND AND AND AND AND AND AND AND AND AND | The Lady Tasting Tea: How Statistics Revolu         by David Salsburg (May 1, 2002)         Average Customer Review: ★★★★★ ⓒ (26)         In Stock         List Price: \$18.89         Price: \$13.88         B1 used & new from \$9.00         ot Interested          atery to added the Theory That Would Not Die to your Wish Li | Add to Cart Add to Wisk           | ıtieth Cent    | tury           |              |
|  | 3.            | The Eights Development of Creation                                  | The Eighth Day of Creation: Makers of the Rk<br>by Horace Freeland Judson (November 1, 1996)<br>Average Customer Review: ★★★★★★ ⊙ (10)<br>In stock on September 4, 2013<br>List Price: \$56.00<br>Price: \$36.09<br>59 used & new from \$26.95<br>to interested © ★★★★★★ Rate this item   | evolution in Biology, 25th Ar     | ıniversary     | <u>Edition</u> |              |
|  | 4.            |   | The Machinery of Life       by David S. Goodsell (April 28, 2009)       Average Customer Review: ★★★★★★ ⓒ (41)       In Stock       List Price: \$25.00       Price: \$17.49       92 used & new from \$12.00   | Add to Cart 🖉 Add to Wisk         | ı List         |                |              |

**Introduction to Machine Learning** 



# **Recommendation Systems**

| Image: Normality in Linking in Change: One of the set Score of Improvement Last Submit Time in Encoder Change in | .e           | aderboard                                  |              | Display top 20 | leaders.            | Reg<br>Slett   | N E T F<br>Browse   |                     | ations             | Friend | s Que  | eue Bu      | uy DVDs          |      |
|--|--------------|--|--------------|----------------|---------------------|--|---------------------|---------------------|--------------------|--------|--------|-------------|------------------|------|
| Jank         Team Name         Best Score         Improvement Last Submit Time           Image: Resemble         0.8853         10.09         2009-07-26 18-18-22           Ballking and Drice Team         0.8871         9.91         2009-07-26 18-18-28           Cound Price Team         0.8871         9.91         2009-07-26 18-18-28           Cound Price Team         0.8871         9.91         2009-07-26 12-07-26           Cound Price Team         0.8873         9.83         2009-07-26 12-07-26           Design Price Team         0.8873         9.83         2009-07-26 12-07-26           Design Price Team         0.8873         9.83         2009-07-26 12-07-26           Design Price Team         0.8873         9.63         2009-07-26 12-07-26           Design Price Team         0.8873         9.43         2009-07-26 12-07-26           Segmetaria         0.88613         9.47         2009-07-26 17-26  |              |  |              |                |                     |  | Home 0              | Genres              | w Releas           | es Pre | views  | Netflix Top | p 100 C          | Crit |
| In the sessments       0.8653       10.19       2009-07-26 18:56.29         Grand Prize Teamsite Chaom       0.8651       9.91       2009-07-26 18:56.29         Grand Prize Teamsite Chaom       0.8671       9.91       2009-07-26 18:56.29         Grand Prize Teamsite Chaom       0.8671       9.91       2009-07-26 18:56.29         Marce Solutions       0.8679       9.83       2009-07-26 12:57.26         Marce Solutions       0.8690       9.71       2009-07-26 12:57.26         Data       0.6693       9.83       2009-07-26 12:57.26         Data       0.6693       9.83       2009-07-26 12:57.26         Data       0.6693       9.84       2009-07-26 12:57.26         Data       0.6693       9.47       2009-07-26 12:57.26         Data       0.6693       9.47       2009-07-26 12:57.26         Battistic       0.6693       9.47       2009-07-26 12:57.26         Battistic       0.6693       9.48       2009-07-26 12:57.26       10:50.44         1       Battistic       0.6693       9.26       2009-07-26 17:52.16       10:50.44         2       Casa Solutions       0.6693       9.26       2009-07-26 17:52.17       10:50.44         3       Casa Solutions   | lank         | Team Name                                  | Best Score   | S Improvement  | Last Submit Time    |  | Mor                 | Inc Enr             | Van                |        |        |             |                  |      |
| Deliking Linkags         Outse 4         1009         2009-07-26 19:18:22           Grand Prize         RMSE <= 0.8563   | 5            | The Ensemble                               | 0.8553       | 10.10          | 2009-07-25 18:38:22 | Contraction of the local diversion of the loc | MOV                 | es ror              | row                |        |        | 1000        | _                |      |
| Grand Prize - RHSE        0.8571       9.81       2009-07-24 13:07:49         Grand Prize - RHSE        0.8579       9.83       2009-07-26 12:07:25         Prizenatificinaria       0.8579       9.83       2009-07-26 12:07:25         Prizenatificinaria       0.8579       9.83       2009-07-26 12:07:25         Parametricinaria       0.8562       9.60       2009-07-26 12:07:25         Dana       0.8563       9.68       2009-07-26 12:07:25         Dana       0.8561       9.49       2009-07-26 12:07:25         Dana       0.8561       9.47       2009-07-26 12:07:26         Dana       0.8561       9.47       2009-07-26 12:07:26         Prior Visco Prize       0.8563       9.42       2009-07-26 12:07:26       0.000-07-26 10:07:26         Stand Drive       0.8563       9.62       2009-07-21 0:20:44:0       0.8567       0.96650       9.68       2009-07-21 0:20:44:0         Stand Drive       0.8568       9.02       2009-07-21 0:20:44:0       0.8667       <  | 0.0          | Bellikons Magmanal Charts                  | 0.8004       | 10,09          | 2009-07-25 18:18:28 |  |                     | Same and the second | and the particular |        |        |             | boull as refered | 24   |
| Scand Directions       0.8871       9.91       2009-07-24 1307-49         Operan Control Internation (Internation (I   | Gran         | <u>d Prize</u> - RMSE <= 0.8563            |              |                |                     | add SnSyster   | Randy, the          | e following movi    | es were            | N      | Caed.  |             |                  |      |
| Cuesse Schriftbane and Vasterie 1       0.8773       9.89       2000-07-25 20.8523         Personaliticitana       0.8779       9.83       2000-07-26 12.5025         Personaliticitana       0.8879       9.83       2000-07-26 12.5025         Personaliticitana       0.8879       9.83       2000-07-26 12.5025         Dans       0.8800       9.71       2000-07-26 12.5725         Dans       0.8801       9.49       2000-07-26 12.5725         Dans       0.8811       9.49       2000-07-26 12.520.552         Dans       0.8813       9.47       2000-07-26 10.200-57         Personalition       0.8813       9.47       2000-07-21 0.200-47         Proprocess Dxike T0005       0.8053       9.26       2000-07-21 0.200-47         Statistion       9.47       2000-07-21 0.200-47       2000-07-21 0.200-47         Statistion       9.47       2000-07-21 0.200-47       2000-07-21 0.200-47         Statistion       9.42       2000-07-21 0.200-47       2000-07-21 0.200-47         Statistion       9.42       2000-07-21 0.200-47       2000-07-21 0.200-47         Statistion       9.46       2000-07-21 0.200-47       2000-07-21 0.200-47         Statistion       9.46       2000-07-21 0.200-47       2000-07  |              | <u>Grand Prize Team</u>                    | 0.8571       | 9.91           | 2009-07-24 13:07:49 | And and determined and the set   | chosen ba           | sed on your inte    | resting            | Allar  | Y      | ou r        | eall             | 1    |
| Manufaling       0.6579       9.83       2009-07-26 (2-48-53)         PreventionEntrans       0.6563       9.58       2009-07-26 (2-28-72.55)       PreventionEntrans       0.6563       9.58       2009-07-26 (2-27.25)         Deals       OperationEntrans       0.85613       9.47       2009-07-26 (2-28.72.55)       PreventionEntrans       0.85613       9.47       2009-07-26 (2-28.42.55)       PreventionEntrans       PreventionEntrans       0.85613       9.47       2009-07-21 (2-04.440         Statistical       0.85613       9.47       2009-07-25 (15-28.34)       2009-07-21 (2-04.440       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       2009-07-25 (15-28.34)       20   | 0.0.0        | Opena Solutions, and Handelay, United      | 0.8573       | 9.89           | 2009-07-25 20:05:52 | Construction and the second se | Carnivale           | Season 1            |                    | G      |        | lika        | 1 :6             | 8    |
| Presentationary       0.8882       9.80       2009-07-26 12:0.983         Pailisor in Eligichess       0.8890       9.71       2009-07-26 12:0.983       2009-07-26 12:0.983         Dase_       0.8891       9.49       2009-07-26 18:0.208       2009-07-26 18:0.208         Destingent in Eligichess       0.8811       9.49       2009-07-26 17:19:11         Destingent in Eligichess       0.8813       9.47       2009-07-26 12:20:652       2009-07-26 12:0:0:428         Processor Environ       0.8813       9.47       2009-07-26 12:0:0:428       2009-07-26 12:0:0:428         Processor Environ       0.8813       9.47       2009-07-26 12:0:0:428       2009-07-26 12:0:0:428         Statistism       0.8853       9.26       2009-07-26 12:0:2:0:428       Statistism       Statistism         Statistism       0.8853       9.26       2009-07-28 10:0:448       Statistism       Statistism       Statistism         Statistism       0.8855       9.02       2009-07-28 10:0:448       Statistism       Statistism       Statistism         Statistism       0.8855       9.02       2009-07-28 10:0:44       Statistism       Statistism       Statistism         Statistism       0.8855       9.02       2009-07-28 10:0:44       Statistism       Statistism  |              | <u>Vandelavindusities i</u>                | 0.8579       | 9.83           | 2009-07-26 02:49:53 | heffitrissank) (   | Fahrenhei           | 19/11               |                    | 1000   |        |             |                  |      |
| Bellikar in EligChassa       0.6580       9.71       2009-07-26 12.87.26         Data       0.6503       9.68       2009-07-26 12.87.26       Control of the big One         Data       0.6612       9.48       2009-07-26 17.18.13       Control of the big One       Control of the big One         Data       0.6612       9.48       2009-07-26 17.18.13       Control of the big One       Control of the big One         Dischassa       0.6613       9.47       2009-07-26 17.18.14       Control of the big One       Control of the big One       Control of the big One         Prosenast Priva       0.6613       9.47       2009-07-26 17.18.14       Control of the big One       Control of t   | 0            | PranmaticThean:                            | 0.8582       | 9.80           | 2009-07-12 15:09:53 |  | 1.10                |                     |                    | upas-6 | Now oy | for         | just \$5.        | 99   |
| Data       0.6803       9.68       2009-07-24 17:18:43         Operations       0.8811       9.49       2009-07-28 18:02:08         De BallSon       0.8612       9.48       2009-07-28 17:19:11         Biblichuma       0.8613       9.47       2009-07-28 17:19:11         Biblichuma       0.8613       9.47       2009-07-24 20:08:48         Propresso Prive 2006 - River 2005 - Witnellog Tozone BallKov in BigChaos       0.8613       9.47         So Sa       0.8614       9.28       2009-07-21 02:04:40         Attractiona       0.6653       9.28       2009-07-21 02:04:40         Attractiona       0.8654       9.14       2009-07-25 17:42:38         So Sa       0.8644       9.14       2009-07-25 17:42:86         So Carlo Carmonizationa       0.8656       9.02       2009-07-25 16:80:34         So Sa       0.8656       9.02       2009-07-25 16:80:34         So Carlo Carmonizationa       0.8656       9.02       2009-07-25 16:80:34         So Stationa       0.8656       9.02       2009-07-25 16:80:34         So Stationa       0.8656       9.02       2009-07-25 16:80:34         So Stationa       0.8656       9.02       2009-07-25 16:80:34         So Stationa <td< th=""><th>000</th><th>Beliker in BigChaos</th><th>0.8590</th><th>9.71</th><th>2009-07-26 12:57:25</th><th>THOUDS</th><th></th><th>The Big (</th><th><u>Dne</u></th><th></th><th></th><th></th><th>-</th><th></th></td<>   | 000          | Beliker in BigChaos                        | 0.8590       | 9.71           | 2009-07-26 12:57:25 | THOUDS   |                     | The Big (           | <u>Dne</u>         |        |        |             | -                |      |
| Conserve Schubbons       0.8611       9.43       2009-07-28 18:02:08         D       BallKar       0.8612       9.43       2009-07-28 18:02:08         D       BallKar       0.8613       9.47       2009-07-28 18:02:08         Progression Prive 20059 - NMGE = 0.3516 - Withming Vestmer BellKas in Blg/Chass       0.8613       9.47       2009-07-28 18:08:14         S       Sampling       0.8613       9.47       2009-07-28 18:08:14       0.00         S       Sampling       0.8613       9.47       2009-07-28 10:08:14         S       Sampling       0.8613       9.47       2009-07-28 10:08:14         S       Sampling       0.8653       9.25       2009-07-28 10:08:14         S       Coma       0.8654       9.25       2009-07-28 10:08:12         S       Coma       0.8656       9.02       2009-07-28 10:08:12         S       Coma (Sampling)       0.8656       9.02       2009-07-28 10:08:14         S       Coma (Sampling)       0.8658       9.02       2009-07-28 10:08:14         S       Coma (Sampling)       0.8658       9.02       2009-07-28 10:08:14         S       Coma (Sampling)       0.8658       9.02       2009-07-28 10:08:14         S  | 0.00         | Date_                                      | 0.8603       | 9.58           | 2009-07-24 17:18:43 | ank) (   |                     | 0 * *               | <b>*</b> \$\$\$    |        | as low |             | utes             |      |
| 0       Balliar       0.8512       9.48       2009-07-26 17:19:11         1       Balliar       0.8513       9.47       2009-06-23 25:06:22         2       Faadure       0.8513       9.47       2009-07-24 20:08:45         Procenses Prive 2005 - BMGE = 0.95016 - Withening Teamer Ballikor In BligCheas       0.8533       9.26       2009-07-21 02:04:00         3       28 martising       0.8534       9.25       2009-07-21 02:04:00         4       Grank       0.8553       9.26       2009-07-25 17:42:38         5       Casa       0.8565       9.02       2009-07-25 17:42:30         6       Liberopin Ball       0.8565       9.02       2009-07-25 17:42:30         7       Liber a case in a canavae       0.8565       9.02       2009-07-25 18:00:54         8       Liberopin Ball       0.8565       9.02       2009-07-25 18:00:54         9       2009-07-25 18:00:54       9.00       2009-07-25 18:00:54         9       Liberopin Ball       0.85658       9.00       2009-07-25 18:00:54         9       2009-07-25 18:00:54       9.09       2009-07-25 18:00:54       9.06         9       2009-07-25 18:00:54       9.09       2009-07-26 16:50:54       9.06         9  | 0            | Opena Solutions                            | 0.8611       | 9.49           | 2009-07-26 18:02:08 | A CONTRACTOR OF A CONTRACTOR O | - E7E               | er su               | m                  |        | 1      |             | 1                |      |
| 1       BitsChram       0.8613       9.47       2009-06-23 23:06:52         2       Fasting       0.8613       9.47       2009-07-24 20:08:48         Propression Prive 20005 - BMACE = 0.25016 - Withming Teams: Boolik to: In BligCheas       0       6000-07-21 02:04:40         3       Attantional       0.8653       9.26       2009-07-21 02:04:40         4       Gramba       0.8654       9.25       2009-07-21 02:04:40         5       Casa       0.8654       9.26       2009-07-21 02:04:40         5       Casa       0.8654       9.17       2009-07-25 17:42:38         5       Invisible Ideas       0.8656       9.02       2009-07-25 17:42:38         6       LDambia Bu       0.8656       9.02       2009-07-25 18:00:54         6       LDambia Bu       0.8656       9.02       2009-07-25 18:00:54         6       LDambia Bu       0.86559       8.99       2009-07-25 18:00:54         6       LDambia Bu       0.86559       8.99       2009-07-11 09:41:54         6       LDambia Su       0.86559       8.99       2009-07-25 18:00:54         6       LDambia Su       0.8658       9.00       2009-07-21 10:20:44         Carant Data       Image Sumba  | 0            | BellKor                                    | 0.8612       | 9.48           | 2009-07-26 17:19:11 | POAT CALL  | -                   |                     | 10/                |        |        | 2           | Original         | art  |
| 2       Fasts2       0.8513       9.47       2009-07-24 200848         Propresse Price 2006 - BitGE = 0.5016 - Winning Teams: Bolliker in BigChaos       0       0       0         3       zianztiening       0.8553       9.26       2009-07-21 02:04:40         4       Grambs       0.8554       9.25       2009-07-25 15:58:34         5       Crait       0.8544       9.17       2009-07-26 15:58:34         5       Crait       0.8544       9.17       2009-07-26 15:58:34         5       Dista Gaunitase       0.8565       9.08       2009-07-26 15:58:34         6       Linking Gaunitase       0.8565       9.08       2009-07-20 03:26:12         7       Justa Gaunitasei       0.8565       9.00       2009-07-20 18:00:54         8       Linking Gaunitasei       0.8565       9.00       2009-07-20 18:00:54         9       Linking Gaunitasei       0.8565       9.00       2009-07-20 18:00:54       0         9       Linking Gaunitasei       0.8565       9.00       2009-07-20 18:00:54       0         9       Linking Gaunitasei       0.8565       9.00       2009-07-20 18:00:54       0         9       Linking Gaunitasei       0.8655       9.00       2009-07-20  | 1            | <u>BigChana</u>                            | 0.8613       | 9.47           | 2009-06-23 23:06:52 | FEREnk) (ssmank):  |                     | -                   | ae                 |        |        |             | originar         |      |
| Brownesse Prive 2009 - BHGE = 0.8616 - Witneling Team: Bellikes in BlgChass         3       genzileni<br>4       0.8653       9.26       2009-07-21 02:04:00         4       Browne       0.8654       9.25       2009-07-25 15:58:34         5       Css       0.8644       9.14       2009-07-26 15:58:34         6       Invisible Ideas       0.8656       9.08       2009-07-21 12:01:42         7       Just a currence       0.8656       9.08       2009-07-22 14:10:42         8       genzit Carmichassi       0.8656       9.02       2009-07-22 14:10:42         9       genzit Carmichassi       0.8656       9.02       2009-07-22 14:10:42         9       asmetrill       0.8658       8.00       2009-07-21 16:00:54         9       asmetrill       0.8659       8.99       2009-07-21 16:00:54       and Cell Salar         9       asmetrill       0.8658       8.00       2009-07-21 10:01:16       and Cell Salar         1       1  | 2            | Feeds2                                     | 0.8613       | 9.47           | 2009-07-24 20:06:46 | (Shideenefit):   |                     |                     |                    |        |        |             |                  | 2    |
| 3       x8mx018min       0.8633       9.26       2009-07-21 02:04:40         4       Stratt       0.8634       9.25       2009-07-25 15:58:34         5       Creas       0.8642       9.17       2009-07-25 15:58:34         5       Creas       0.8644       9.14       2009-07-25 15:58:34         5       Invisible Ideas       0.8656       9.08       2009-07-25 15:00:54         5       Creas       0.8656       9.02       2009-07-22 14:10:42         5       Crease Cammichases       0.8656       9.02       2009-07-22 14:10:42         6       asmethill       0.8656       9.02       2009-07-22 14:10:42         6       asmethill       0.8656       9.02       2009-07-21 10:92:12       10:2:0:2:0:0:0:41         7       Jamte Cammichases       0.8656       9.02       2009-07-22 14:10:12       Create Cammichases       0.8656       9.02       2009-07-21 14:10:12       Create Cammichases       0.8655       8.99       2009-04-16 00:29:35       reate Cammichases       0.8655       8.99       2009-04-16 00:29:35       reate Cammichases       1       ?       3       5         Phoine 2002 - Bubble  |              | ress Prize 2008 - RMSE = 0.5616 - 1        |              |                | 1905                | Carterance, Sanathan T   | dRene               | 1                   |                    | OII    |        |             | Gr               | 2    |
| 4       Arank       0.8634       9.25       2009-07-26 15:86:34         5       Chan       0.8642       9.17       2009-07-26 17:42:38         5       Invisible Ideas       0.8644       9.14       2009-07-26 17:42:38         5       Chan a caravae       0.8656       9.08       2009-07-22 14:10:42         6       Chan be caravae       0.8656       9.02       2009-07-22 14:10:42         6       Chan be caravae       0.8656       9.02       2009-07-22 14:10:42         6       Arank       0.8656       9.02       2009-07-25 18:00:42         6       Arank       0.8656       9.02       2009-07-25 18:00:42         6       Arank       0.8656       9.02       2009-07-25 18:00:42         7       Junta curve       0.8656       9.02       2009-07-25 18:00:42         8       Chance Caravae       0.8656       9.02       2009-07-25 18:00:42         9       J.Darta curve       0.8656       9.02       2009-07-25 18:00:42         9       J.Darta curve       0.8656       9.02       2009-07-25 18:00:42         Phoree core curve       0.8656       9.02       2009-04-16 05:28:35         Phoree core curve       1       ?       3   | 3            | xiangliang                                 | 0.8633       | 9.26           | 2009-07-21 02:04:40 | CE COLL  | + 345               | N' ~                |                    |        |        |             |                  |      |
| 5       Csa       0.8642       9.17       2009-07-25 17:42:38         5       Invisible Ideas       0.8644       9.14       2009-07-20 03:26:12         7       Just a curvina a carmene       0.8655       9.08       2009-07-22 14:10:42         2       Carate Carmichasei       0.8656       9.02       2009-07-25 17:42:38         3       LDaronia Ba       0.8656       9.02       2009-07-25 18:00:54         2       2009-07-25 18:00:54       2009-07-25 18:00:54       2009-07-25 18:00:54         3       J.Daronia Ba       0.8656       9.02       2009-07-25 18:00:54         2       2009-07-25 18:00:54       2009-07-25 18:00:54       2009-07-25 18:00:54         2       Description Ba       0.8656       9.02       2009-07-25 18:00:54         Procenta Ba       0.8659       9.09       2009-04-16 06:29:35       2009-04-16 06:29:35         Procenta Carmedichi scolares on guild: colleses - sites       Image: Carmedichi scolares on guild: colleses - sites       Image: Carmedichi scolares on guild: colleses - sites         Charmedichi scolares on guild: colleses - sites       Image: Carmedichi scolares on guild: colleses - sites       Image: Carmedichi scolares on guild: colleses - sites         Charmedichi scolares on guild: colleses - sites       Image: Carmedichi scolares on guild: colleses - sites  | 4            | <u>Gravity</u>                             | 0.8634       | 9.25           | 2009-07-26 15:58:34 | startu   | lland               |                     |                    |        |        | 5           | and Sa           | RUN  |
| 5       Invisible Ideas       0.8544       9.14       2009-07-20 03.2612         7       Just & divine a camarae       0.8550       9.08       2009-07-22 14:10:42         2       Carle Carmichae       0.8556       9.02       2009-07-25 16:00:54         2       LDannia Ru       0.8556       9.02       2009-07-25 16:00:54         2       acmetrili       0.8559       9.09       2009-07-16 06:29:35         Procurses Price 2007       NMSE = 0.9712 - Winning Team: BortBall       2009-04-16 06:29:35         Chronewestich scores on guitz solazes - BINSE = 0.9514       Control Canada       Image: Canada Canada Canada  | 5            | Ces  | 0.8642       | 9.17           | 2009-07-25 17:42:38 | sistance   | BD->                | Season 2            |                    |        | a.     |             | 0                |      |
| 7       Just & quoin & garage       0.8650       9.08       2009-07-22 14:10:42         8       Crant Cannotshage       0.8656       9.02       2009-07-25 16:00:64         9       J.Dambia Ru       0.8656       9.00       2009-07-25 16:00:64         9       J.Dambia Ru       0.8656       9.00       2009-07-25 16:00:64         9       J.Dambia Ru       0.8659       8.99       2009-04-16 06:29:35         Procuresos Prices 2002 - NMSE = 0.951/2       Winning Teamu: MontBell       This state of the stat  | 6            | Invisible Ideas                            | 0.8644       | 9.14           | 2009-07-20 03:26:12 | cance, SofransactitO   | smank)              | Disc Seri           | e                  |        | r      |             | T                |      |
| 3       Carnet Carmetrines       0.8656       9.02       2009-07-25 18:00:54         9       LDemoin Bu       0.8558       9.00       2009-07-25 18:00:54         9       LDemoin Bu       0.8559       9.00       2009-07-15 06:29:35         Procress Price 2002/ - MinStE = 0.5712 - Winning Yeam: Konticul       2009-04-16 06:29:35         Cincernatich score on guits subset - NMSE = 0.5514       Control of the score on guits subset - NMSE = 0.5514   | 7            | Just a que in a gamae                      | 0.8650       | 9.08           | 2009-07-22 14:10:42 | Constantsistance-set   | 11/100              |                     | ÷.                 |        |        |             |                  |      |
| asymptotic Bit 0.8558       0.00 2009-03-11 09:41:54 2009-04-18 06:29:35         bit scores in quite subsets - EWSE = 5.953.4       2009-04-18 06:29:35         Cincernation Scores on quite subsets - EWSE = 5.953.4       Image: Constraint of the scores on quite subsets - EWSE = 5.953.4  | 8            | Crate Carrichael                           | 0.8656       | 9.02           | 2009-07-25 16:00:54 | Since 65   | 5Banr               |                     |                    |        |        |             |                  |      |
| 0 i somehili       0.8559 i 8.99 i 2009-04-16 05-28:35         Progresse Prize 2002 - Striket = 0.62/12 - Witheling Yearm: Kortbell         Chrosmatich scores on guite subjects - Striket = 0.95/14   | 9            | <u>, I Dennie Bu</u>                       | 0.8658       | 9.00           | 2009-03-11 09:41:54 | FartDat  | aAs 🗖               |                     |                    |        | - 115  | ers         |                  | 4    |
| Progress Price 2007 - NMSE = 0.8712 - Winning Team: RorBell         Chromatich come on quic subast - RMSE = 0.9514         Chromatich come on quic subast - RMSE = 0.9514  | 0            | <u>armehill</u>                            | 0.8659       | 8.99           | 2009-04-16 06:29:35 | Sale Const Of seton, Sasan   | op All Intur        |                     |                    |        | uc     |             |                  | r"   |
| Chrematch score on quiz subset - BNSE = 5.95%  | Prog         | <u>ness Prize 2007</u> - RMSE = 0.8712 - 1 | Winning Team | : KorBell      |                     | BD->Reco   | Ve Chies v<br>stbow | •                   | 1                  |        | ?      | 3           | 5                |      |
| startpala  | <u>Ciner</u> | natch score on quiz subset - 845E          | = 0.9514     |                |                     | 18/11/1. 15  |                     | moules              | 2                  | 4      |        |             |                  | t    |
| Startpial  |              |  |              |                |                     | (\$nIdBer  | Roder #1            | novies              | 1                  |        | -      |             |                  |      |
|  |              |  |              |                |                     | startData  | -                   |                     |                    |        |        |             |                  | t    |

#### Machine learning competition with a \$1 million prize

#### **Introduction to Machine Learning**

#### Inas A. Yassine

2



# Clustering

**Discovering Structures in Data** 

**Introduction to Machine Learning** 

# **Clustering Images**





Inas A. Yassine

Cairo University



# **Clustering Web Search Results**



#### Introduction to Machine Learning



# What is Machine Learning

(Definition)

**Introduction to Machine Learning** 

# Why "Learn"?



- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to "learn" to calculate payroll
- It is about seeking a predictive and/or executable understanding of natural/

artificial subjects, phenomena, or activities from ...

- 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)



# What is Machine Learning(ML)?

• Arthur Samuel: "It is a science that gives the computer the ability to learn without the need to write a program".



## What We Talk About When We Talk About "Learning"



- Learning general models from a 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



Is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data

#### **Applications:**

- 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

• ...



# Supervised Learning : find f

- Given: Training set  $\{(x_i, y_i) | i = 1 ... N\}$
- 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 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  $f: X \rightarrow Y$ 
  - $X = \{0, 1\}^4$
  - $Y = \{0, 1\}$
- Question1: How should we pick the hypothesis space, the set of possible functions f?
- Question2: 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<sup>16</sup> possible hypotheses
- 2<sup>9</sup> 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     | 7 |

#### Dataset:

| Example | x <sub>1</sub> | $x_2$ | X3 | x <sub>1</sub> | 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 |



# UnSupervised Learning : find groups

- Given: Training set  $\{(x_i) | i = 1 ... N\}$
- Find: Learning "what normally happens" →No output
- Clustering: Grouping similar instances
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs



#### **Introduction to Machine Learning**

# **Reinforcement Learning**



#### • Topics:

- Policies: what actions(Reward, punishement) should an agent take in a particular situation
- Utility estimation: how good is a state ( $\rightarrow$ used by policy)
- No supervised output but delayed reward
- Credit assignment problem (what was responsible for the outcome)
- Applications:
  - Game playing
  - Robot in a maze
  - Multiple agents, partial observability, ...

#### **Introduction to Machine Learning**



# **Medical Informatics**

**Introduction to Machine Learning** 

### **Medical informatics**



Medical informatics is the application of computers, communications and information technology and systems to all fields of medicine \*

- medical care
- medical education
- medical research.

\*MF Collen, MEDINFO '80, Tokyo

#### **Introduction to Machine Learning**

### What is medical informatics?



Medical informatics is the branch of science concerned with the use of computers and communication technology to acquire, store, analyze, communicate, and display medical information and knowledge to facilitate understanding and improve the accuracy, timeliness, and reliability of decision-making.\*

\*Warner, Sorenson and Bouhaddou, Knowledge Engineering in Health Informatics, 1997

**Introduction to Machine Learning** 

# **Clinical decision making**



- Making sound clinical decisions requires:
  - right information, right time, right format
- Clinicians face a surplus of information
  - ambiguous, incomplete, or poorly organized
- Rising tide of information
  - Expanding knowledge sources
    - 40K new biomedical articles per month
    - Publicly accessible online health info
    - Hundreds of pictures per scan for one patient

# Clinical decision making: What is the problem?



- Man is an imperfect data processor
  - We are sensitive to the *quantity* and *organization* of information
- Army officers and pilots commit 'fatal errors' when given too many, too few, or poorly organized data
- The same is true for clinicians who 'watch' for events
- Clinicians are particularly susceptible to errors of omission



# Clinical decision making: What is the problem?

- Humans are "non-perfectable" data processors
  - Better performance requires more time to process
  - Irony
    - Clinicians increasingly face productivity expectations
    - Clinicians face increasing administrative tasks

### What is medical imaging (MI)?



The study of medical imaging is concerned with the interaction of all forms of radiation with tissue and

the development of appropriate technology to extract clinically useful information (usually displayed in an image format) from observation of this technology.

#### **Sources of Images:**

- Structural/anatomical information (CT, MRI, US) within each elemental volume, tissue-differentiating properties are measured.
- Information about function (PET, SPECT, fMRI).

Introduction to Machine Learning

# The imaging "chain"





# **Computer-Aided Diagnosis**



• **Computed Aided Diagnosis (CAD)** is diagnosis made by a radiologist when the output of computerized image analysis methods has been incorporated into his or her medical decision-making process.

- CAD may be interpreted broadly to incorporate both
  - the detection of the abnormality task and
  - the classification task: likelihood that the abnormality represents a malignancy

# Motivation for CAD systems



becoming overwhelming for human vision and the overload of image data for interpretation may result in oversight errors.

The amount of image data acquired during a CT scan is

#### **Computed Aided Diagnosis for:**

- Breast Cancer
  - Mammograhy Images
- Lung Cancer
  - A thoracic CT scan generates about 240 section images for radiologists to interpret.



# Content Based Medical Image Retrieval

**Introduction to Machine Learning** 

# Content-based medical image retrieva (CBMS) systems

**Definition of Content-based Image Retrieval:** 

**Content-based image retrieval** is a technique for retrieving images on the basis of automatically derived image features such as texture and shape.

#### **Applications of Content-based Image Retrieval:**

- Teaching
- Research
- Diagnosis
- PACS and Electronic Patient Records



#### Inas A. Yassine

Cairo University

**Introduction to Machine Learning** 

# **CBIR** as a Diagnosis Aid



An image retrieval system can help when the diagnosis depends strongly on direct visual properties of images in the context of evidence-based medicine or case-based reasoning



Introduction to Machine Learning



# **CBIR as a Teaching Tool**

An image retrieval system will allow students/teachers to browse available data themselves in an easy and straightforward fashion by clicking on "show me similar images".

#### Advantages:

- stimulate self-learning and a comparison of similar cases

- find optimal cases for teaching Teaching files:

- Casimage: <u>http://www.casimage.com</u>
- myPACS: <u>http://www.mypacs.net</u>

|  |   |  | - airo                                | Universit,                 |
|--|---|--|---------------------------------------|----------------------------|
| MyPACS<br>Teaching File  | .net<br>Management System   |  | Search in: 🔍 All Case                 | C My Cases Advanced Search |
| Home About   | Support Solutions Ca  | ases   |                                       |                            |
|  | Welco   | me Log in Sign-i   | up                                    |                            |
| MyPACS Log In Login name: Password: Sign in Sign up for your free account Case of the Week APPENDICITIS by James Patrick, Resident, Medical College of Ohio, USA | Create Your Own<br>Teaching Files<br>Join radiologists from 4<br>institutions in 75 count<br>are using MyPACS to c<br>cases online. Users hav<br>contributed <u>5000 radio</u><br><u>teaching files contain</u><br><u>20,000 images</u> , and r<br>are being added every<br>you need Is a free accor<br>your web browser, and<br>start creating cases rig<br>This is a free service to<br>international communi<br>in part by the National<br>of Health. We also offer<br><u>enterprise teaching file</u> | 400<br>tries who<br>reate their<br>ve<br>ology<br>ning<br>new cases<br>day. All<br>ount and<br>lyou can<br>ht now.<br>o the<br>ty, funded<br>I nstitutes<br>r custom<br>a solutions. |                                       |                            |
| Notices IV Provide Texas   |   | Shared Cases   |                                       |                            |
| R R  | Cranium and<br>Contents<br>Face and Neck<br>Spine and Peripheral<br>Nervous System<br>Skeletal System   | Heart<br>Lung<br>Gastrointestinal<br>(GI)<br>Genitourinary<br>(GU)   | Vascular/Lymphatic<br>Breast<br>Other |                            |
|  | Highlighted Featu   | reci   |                                       |                            |
|  |   |  |                                       |                            |

# **CBIR as a Research Tool**



Image retrieval systems can be used:

to complement text-based retrieval methods

 for visual knowledge management where by the images and associated textual data can be analyzed together

 multimedia data mining can be applied to learn the unknown links between visual features and diagnosis or other patient information

 for quality control to find images that might have been misclassified

# In Summary ...



- Machine Learning is a Knowledge Discovery Process
- 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 due to
  - Big data
  - Improved machine learning algorithms Faster computers
  - Good open-source software
- Data Mining Tasks
  - classification, clustering, regression, ...

# **Course Content**



- Introduction: Machine Learning for health informatics
- Linear prediction: Regression
- Learning strategies for basic classification algorithms: Linear classification, logistic regression
- Probabilistic models ( Bayes Theory, Decision trees)
- Learning strategies for more advanced classification algorithms (SVM, ...)
- Dimensionality Reduction (Principal component analysis )
- Subspace Clustering
- Neural Networks
- Reinforcement Learning
- Recommender Systems
- Overfitting, underfitting, and the generalization gap
- Machine Learning Algorithms testing and evaluation

# Cairo University

# **Grading Criteria**

| Item                  | <b>Course Grade Contribution</b> |
|-----------------------|----------------------------------|
| Final Exam            | 70%                              |
| Final Project         | 8%                               |
| Assignments/ Homework | 15%                              |
| Midterm Exam          | 7%                               |

# **Cheating Description**

#### What is cheating?

- Sharing code: by copying, retyping, looking at, or supplying a file
- Describing: verbal description of code from one person to another.
- Coaching: helping your friend to write a line by line code
- Searching the Web for solutions
- Copying code from a previous course or online solution
- You are only allowed to use code we supply
- What is NOT cheating?
  - Explaining how to use systems or tools
  - Helping others with high-level design issues
- Penalty for cheating
  - Zero in Assignments/ Exam





# In Class/ Assignment General Code

Electronic communications: forbidden

- No email, instant messaging, cell phone calls, etc
- Presence in lectures: highly recommended
- No recordings of ANY KIND
- No side talks



Thank You ...

**Introduction to Machine Learning**