Decision Trees & CBC Intro – Lecture Overview

- Problem Representation using a Decision Tree

- ID3 algorithm

- The problem of overfitting

- Research on affective computing, natural HCI, and ambient intelligence

- Facial expressions and Emotions

➢ Overview of the CBC

➢ Group forming

CBC - Goal

- Hands-on experience in implementing and testing basic machine learning techniques

- Work with other team members

➢ CBC = Computer-Based Coursework

Group Forming

- Students will be divided in groups of 4

- Simply fill in the **excel form** with the following information (for each group member):

  - First Name, Last Name, Student login, Email, Course, CID

- You can find the excel form on http://ibug.doc.ic.ac.uk/courses,

  http://ibug.doc.ic.ac.uk/media/uploads/documents/ml-cbc-groupform.xls

COMPUTER BASED COURSEWORK (CBC)

CBC contents:

The CBC is designed to build on lectures by teaching students how to apply ML techniques about which they have been lectured to real-world problems.

The CBC will consist of three assignments. All three assignments will focus on emotion recognition from data on displayed facial expression using decision trees, neural networks, and case-based reasoning. The last assignment will also focus on evaluating (by means of paired t-tests) which of these ML techniques is more suitable for the problem in question in the case of clean data and in the case of noisy data.

CBC assessment:

Assessment of the CBC work will be conducted based upon the following:

- the quality of the delivered code as measured by the clarity, effectiveness and efficacy of the delivered code when tested on real (previously unseen) data,

- the quality of the delivered reports for each of the CBC assignments as measured by the correctness, depth and breadth of the provided discussion on the evaluation of the performance of the developed ML systems for emotion recognition,

- individual involvement and contribution to the group's results (to be judged based upon a final interview with each of the groups).

CBC data and tools:

You can download all the required datasets and the software tools that you need to use in one zip file here.

CBC contact:

All Teaching Helpers can be contacted via one email address. If you wish to contact a specific TH, specify the TH's name in the subject of your email.

Group formation:

Please email us this form through this e-mail address to enrol in CBC.

FURTHER READING: LECTURE SLIDES

Group Forming

- Email the **excel form** by noon **Wednesday January 25th**
  machinelearningtas@gmail.com

- You will be assigned a TH who will email you informing you about your group number and confirming the members of your group by the end of the week

Group Forming

- 4 members per group, this is a hard limit!

   - In other words, groups of 5, 6, 7, etc are not allowed!!

- If you cannot form a team with 4 members then just email us the above information and we will assign you to a team.

   - Case 1 member: We will add you to a group of 2-3 members
   - Case 2 members: 1-2 more members will be added to your group
   - Case 3 members: 1 member may be added to your group

- We sometimes allow groups of 3 (depends on the number of students who take the course) but generally you should expect that a 4th member will be added by us.

- Once the groups are formed you cannot change groups

External Students

- Follow instructions provided by the department

Tutorial Helpers

- *A Tutorial Helper (TH) will be assigned to each group*

  - *Eftychia Fotiadou*
  - *Markos Georgopoulos*
  - *Jean Kossaifi*
  - *Kritapat Songsri-In*
  - *Yujiang Wang*
  - *Robert Walecki*
  - *Linh Tran*
  - *Mengjiao Wang*
  - *Jie Pu*
  - *Jiankang Deng*
  - *Yiming Lin*
  - *Christopher Bowles*

http://ibug.doc.ic.ac.uk/people
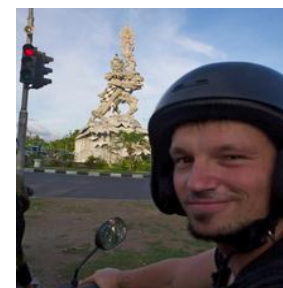
- In the labs you can ask questions to any TH

Communication

➢ Via the website: http://ibug.doc.ic.ac.uk/courses/machine-learning-course-395/

  - CBC Manual (Section Course Material)

  - Provided Matlab files, datasets (Section Data and Tools)

  - Tutorials

COMPUTER BASED COURSEWORK (CBC)

CBC contents:

The CBC is designed to build on lectures by teaching students how to apply ML techniques about which they have been lectured to real-world problems.

The CBC will consist of three assignments. All three assignments will focus on emotion recognition from data on displayed facial expression using decision trees, neural networks, and case-based reasoning. The last assignment will also focus on evaluating (by means of paired t-tests) which of these ML techniques is more suitable for the problem in question in the case of clean data and in the case of noisy data.

CBC assessment:

Assessment of the CBC work will be conducted based upon the following:

- the quality of the delivered code as measured by the clarity, effectiveness and efficacy of the delivered code when tested on real (previously unseen) data,

- the quality of the delivered reports for each of the CBC assignments as measured by the correctness, depth and breadth of the provided discussion on the evaluation of the performance of the developed ML systems for emotion recognition,

- individual involvement and contribution to the group's results (to be judged based upon a final interview with each of the groups).

CBC data and tools:

You can download all the required datasets and the software tools that you need to use in one zip file here.

CBC contact:

All Teaching Helpers can be contacted via one email address. If you wish to contact a specific TH, specify the TH's name in the subject of your email.

Group formation:

Please email us this form through this e-mail address to enrol in CBC.

FURTHER READING: LECTURE SLIDES

Communication

➢ Via email: machinelearningtas@gmail.com

ALWAYS put your group number in the subject line

CBC – Organisation

- Each group must hand in a report of ~4-5 pages (excluding figures) per assignment, including discussion on implementation and answers to questions posed in the manual.

- ONE report per group

- Each group must hand in the code they implemented for each assignment.

- Hand in the code and the reports via CATE.

CBC – Assignment hand in

- Hand in via CATE

  - One group leader per group

  - Each and every group member individually has to confirm that s(he) is part of that particular group, for each and every assignment submission (under the pre-determined group leader) before each assignment submission deadline.

Imperial College
London

CBC – Report + Code marking

- The THs will mark your report and provide feedback

- The THs will test the implemented algorithms using a separate test set (not available to the students).

- We will inform you about the performance of your algorithm on our test set.

- We just want to check that your code runs and your classifier has been properly trained, i.e., it can generalise to unknown data

Lab Schedule

Assisted Labs (THs present to answer questions), starting on January 26th continuing until first week of March (**NO LAB THIS WEDNESDAY**)

- Every Wednesday 9:00-11:00, lab 219

- Every Thursday 16:00 – 18:00, lab 219
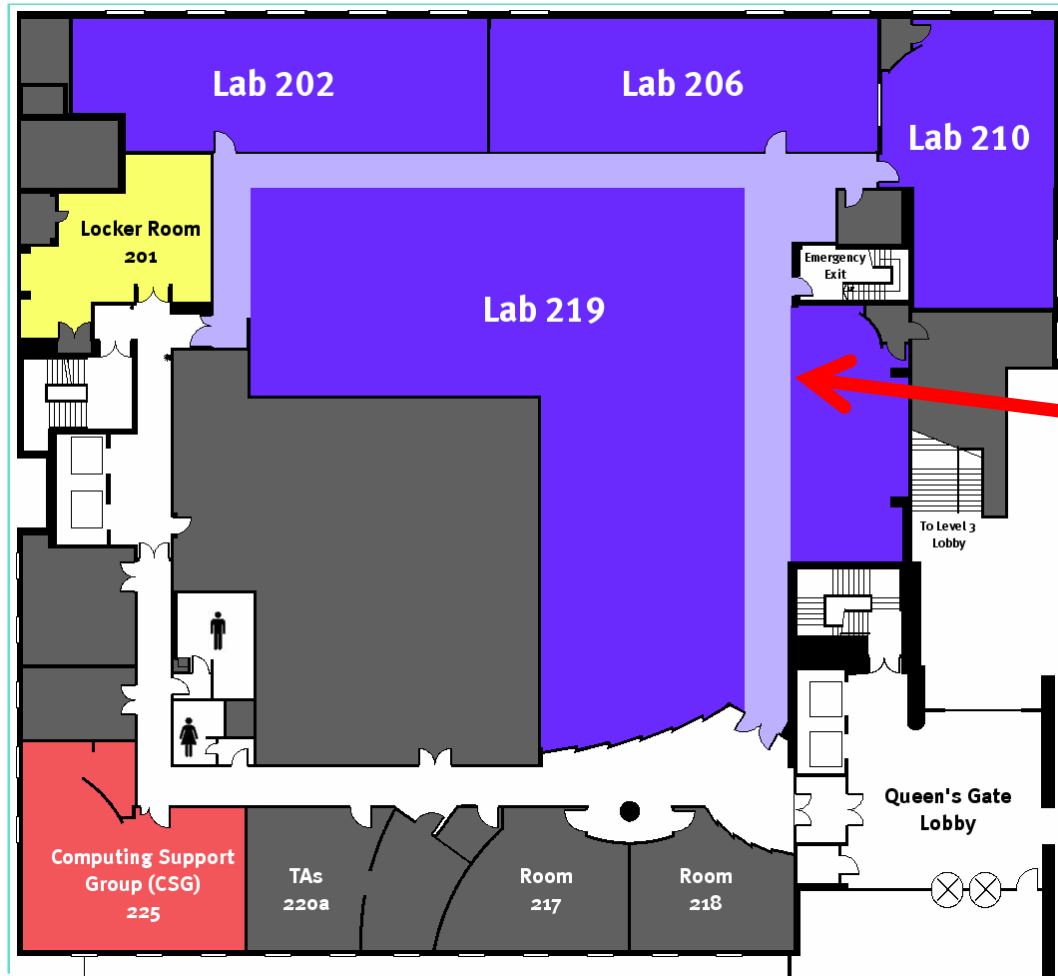
THs will
be here

Deadlines

- Assignment 1: optional (no hand in required)

- Assignment 2: February 13th (Monday) – noon

- Assignment 3: March 6th (Monday) – noon

CBC – Grading

- Every group member is expected to have sufficient contribution to the implementation of every assignment.

- Plagiarism is not allowed!
  Involved groups will be instantly eliminated.

Assignment Grading
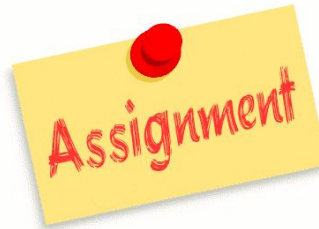
Report Content   *Code*   *Report Quality*
   75%   *15%*   *10%*

Group_grade = 0.75*report_content + 0.15*code + 0.1*report_quality

- *Code Mark:* Depends on the performance on the hidden Test set

  e.g. for Decision Trees = performance on hidden Test set + 15

  (Max Performance in previous years was 85%)

- *Make sure your code runs. If not you lose 30% of the code mark.*

Assignment Grading

Grade1   **Grade2**

CBC_grade = 0.4*Grade1 + 0.6*Grade2

Machine Learning Grade

CBC Grade

33.3%

Exam Grade

66.7%

- *CBC accounts for 33.3% of the final grade for the Machine Learning Course. In other words, final grade = 0.667*exam_grade + 0.333*CBC_grade.*

CBC – Tools

- Training data and useful functions are provided via the course website in a separate .rar file.

- Courseworks have been designed and tested in MATLAB
  - ➢ MATLAB basics (Assignment 1)
  - ➢ Decision Trees (Assignment 2)
  - ➢ Neural Network Toolbox (Assignment 3)

- Students are strongly advised to use the MATLAB help files!

- You can also use Python BUT you should write all the necessary code (we only provide Matlab functions for Decision trees visualisation etc)

Assignment 1 : MATLAB Exercises

- Optional (no hand in required)

- A brief introduction to some basic concepts of MATLAB without assessing students' acquisition, application and integration of this basic knowledge.

- The students, are strongly encouraged to go through all the material, experiment with various functions, and use the MATLAB help files extensively (accessible via the main MATLAB window).

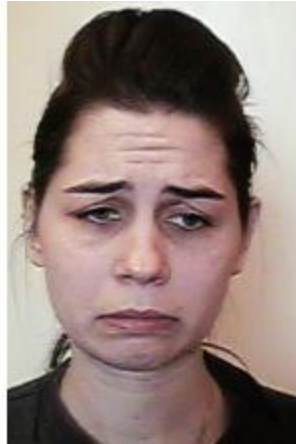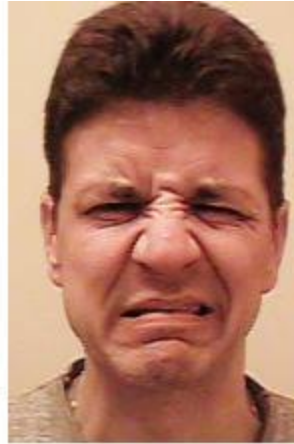# CBC – Emotion Recognition
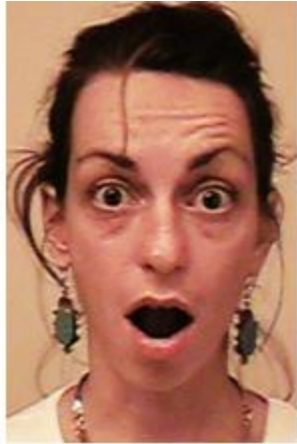
Anger   Surprise   Sadness   Disgust   Fear   Happiness

- Prototypic facial expressions of the **six basic emotions** were introduced by Charles Darwin (1872) and elaborated by Ekman

- These prototypic facial expressions can be described in terms of AUs (e.g., surprise $\leftrightarrow$ AU1 + AU2 + AU5 + AU26 / AU27)

Facial Muscle Actions (Action Units - AUs)

CBC – Emotion Recognition

Emotion	AUs	Emotion	AUs
Happy	{12}	Fear	{1,2,4}
	{6,12}		{1,2,4,5,20,
Sadness	{1,4}		25||26||27}
	{1,4,11||15}		{1,2,4,5,25||26||27}
	{1,4,15,17}		{1,2,4,5}
	{6,15}		{1,2,5,25||26||27}
	{11,17}		{5,20,25||26||27}
	{1}		{5,20}
Surprise	{1,2,5,26||27}		{20}
	{1,2,5}	Anger	{4,5,7,10,22,23,25||26}
	{1,2,26||27}		{4,5,7,10,23,25||26}
	{5,26||27}		{4,5,7,17,23||24}
Disgust	{9||10,17}		{4,5,7,23||24}
	{9||10,16,25||26}		{4,5||7}
	{9||10}		{17,24}

V: $AUs \rightarrow$ *basic-emotions*

V': $\langle a_1, \ldots, a_{45} \rangle \rightarrow [1..6]$

learning algorithms:
- decision trees (ID3)
- Neural Networks
- Case-based Reasoning

evaluating developed systems:
- t-test

Assignments 2-3 : Overview

- Classification Problem

  - Inputs: *x (AU vectors)*

  - Desired Output: *y (Emotion label)*

- Use *x* and *y* to train your learning algorithms
  to discriminate between the 6 classes (emotions)

- Evaluate your algorithms using 10-fold cross validation (decision trees) or the holdout method (training/validation/test sets)

- Write a function $y^{pred} = testLearner(T, x)$, which takes your trained learners T and the features *x* and produces a vector of label predictions y^{pred}

Imperial College London
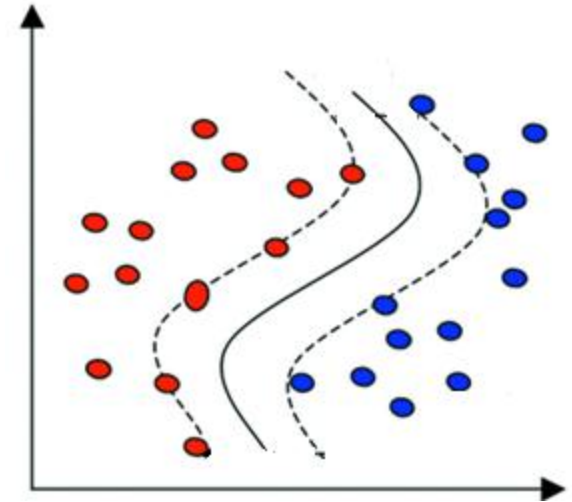
Training – Validation – Test Sets

- Training Set: Used to train the classifiers

- Validation Set: Used to optimise the parameters of the classifiers
  - e.g. number of hidden neurons in neural networks

- Test Set: Used to measure the performance of the classifier

N-fold Cross validation

- Total error estimate: $\hat{e} = \frac{1}{N}\sum_{i=1}^{N} e_i$
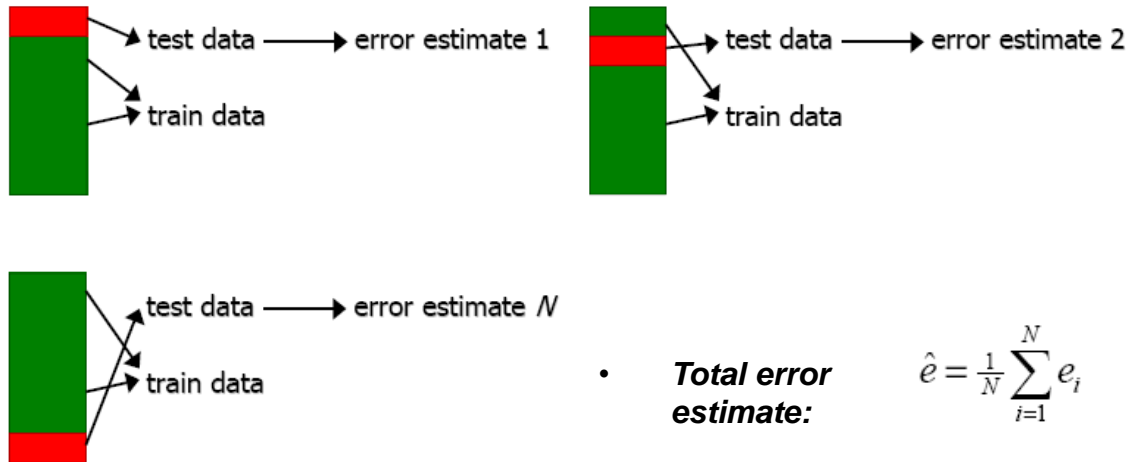
- Initial dataset is partitioned in N folds
- Training + **Validation** set: N - 1 folds, Test set: 1 fold
- This process is repeated N times → N error estimates
- Final error: Average of the N error estimates

Assignment 2 : Decision Trees

- Implement and train a decision tree learning algorithm

- Evaluate your trees using 10-fold cross validation

- Write a function $y^{pred} = testTrees(T, x)$, which takes your trained trees T and the features x and produces a vector of label predictions y^{pred}

- If you use Python you should provide clear instructions how to test your trained trees

- Theoretical / Implementation questions

Assignment 3 : Artificial Neural Networks

- Use the Neural Networks toolbox (MATLAB built-in) to train your networks

- Evaluate your networks on the test set

- Write a function: y^{pred} = testANN(N, x), which takes your trained networks N and produces a vector of label predictions y^{pred} .

- If you use Python you should provide clear instructions how to test your trained networks.

- Theoretical / Implementation questions

CBC – Tips

- Work together!!! Divide the work, but you should all know how you have implemented your algorithms, why you took some specific decision, etc...

- Make sure your code runs!!

- **Make sure you use the new version of the manual (updated on Friday!)**

- If you have any complaints (marking, THs etc) contact the course support leader

Group Forming

- Students will be divided in groups of 4 students

- Simply fill in the **excel form** with the following information (for each group member):

  -Student login, email, First Name, Last Name, Course, CID

- You can find the excel form on http://ibug.doc.ic.ac.uk/courses, Section: Group Forming

  http://ibug.doc.ic.ac.uk/media/uploads/documents/ml-cbc-groupform.xls

- Email the **excel form** by noon **Wednesday January 25th**
  machinelearningtas@gmail.com

- If you cannot form a team with 4 members then just email us the above information and we will assign you to a team.