Today's Schedule

- Topic revision
- Maths formulae to remember
- Methods/derivations to understand
- Exam technique

Maths formulae to remember

- Euclidean distance:
  \[ r_2(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2} \]
- cf. \( \sin(x, y) = \frac{\sum_{i=1}^{D} (x_i - y_i)}{\sqrt{\sum_{i=1}^{D} (x_i - y_i)^2}} \) as a similarity measure
- Pearson correlation coefficient:
  \[ \rho(x, y) = \frac{\sum_{i=1}^{D} (x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y} \]
- Bayes Theorem
  \[ P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)} \]
- K-means clustering:
  \[ P(C_k | x) = \frac{P(x | C_k)P(C_k)}{\sum_k P(x | C_k)P(C_k)} \]
- Logistic sigmoid function:
  \[ y = g(a) = \frac{1}{1 + \exp(-a)} \]
- Softmax activation function (for multiple output nodes):
  \[ y_k = \frac{\exp(a_k)}{\sum_{k=1}^{K} \exp(a_k)} \]
- and basic maths rules (e.g. differentiation)

Maths formulae to remember (cont.)

- Bayes decision rule (cf. MAP decision rule)
  \[ k^* = \arg \max_{k} P(C_k | x) = \arg \max_{k} P(x | C_k)P(C_k) \]
- Naive Bayes for document classification
  \[ P(C_k | X) = \prod_{i=1}^{D} P(w_i | C_k) \]
- Logistic sigmoid function:
  \[ y = g(a) = \frac{1}{1 + \exp(-a)} \]
- Softmax activation function
  \[ y_k = \frac{\exp(a_k)}{\sum_{k=1}^{K} \exp(a_k)} \]
- and basic maths rules (e.g. differentiation)

Methods/derivations to understand (non exhaustive)

- Clustering and classification
- Discriminant functions of Gaussian Bayes classifiers
- Learning as an optimisation problem
  - Maximum likelihood estimation
  - Gradient descent and back propagation algorithm (neural networks) for minimising the sum-of-squares error
- NB: Learning is a difficult problem by nature — generalisation from a limited amount of training samples. Need to assume some structures (constraints).

Machine learning as optimisation problems

- Euclidean-distance based classification
  \[ k^* = \arg \min_{l < k} ||x - r_l|| \]
- K-means clustering
  \[ \min_{\mu_k \in \mathbb{R}^D} \sum_{x \in X} ||x - \mu_k||^2 \]
- Dimensionality reduction to 2D with PCA
  \[ \text{max } \text{Var}(y) = \text{Var}(z), \quad z = \mu + w^T (x - \mu) \]
- Bayes decision rule
  \[ k^* = \arg \max_{k} P(C_k | x) = \arg \max_{k} P(x | C_k)P(C_k) \]
- Maximum likelihood parameter estimation
  \[ \text{max } \ln P(X | \theta) \]
- Least squares error training of neural networks
  \[ \min_{\theta} \sum_{n=1}^{N} ||y_n - \hat{y}_n||^2 \]
Exam revision

Look at lecture notes, slides, tutorials, coursework, and past papers.

Early exam papers: many (useful) multiple choice Qs
- No longer the exam format
- Syllabus has changed slightly

Recent exam papers since 2008/09
- Answer two questions from section A (ADS) and two questions from section B (Learning).
- Closed-book exam.
- Calculators may be used (approved ones only).
- Solutions are available only for 2008/09, 2009/10, 2013/14 (no plans of releasing those of missing years)
- NB: errors in some solutions, e.g. 5 (c) of 2008/09: square root is not taken in computing standard deviations.

Well prepared for the exam of 120 minutes
- 60 minutes/section, 30 minutes/question

Exam revision (cont.)

Don’t overfit!
- Anything that appears in the notes, slides, tutorial sheets, or coursework is examinable, unless marked non-examinable, extra topics, or (*)

Don’t trust unofficial solutions

Inf2b Revision Meeting
- Date: TBC (in late April)
- Send me questions/requests that you want me to discuss at the meeting.

Time in the exam

- Half an hour per question (minus time to pick questions)
- Don’t panic!
- Go for easy marks first
- Don’t spend a long time on any small part
- Don’t spend a long time on any small part
- Don’t scrawl - you might lose marks if the marker cannot read/understand
- Know the standard stuff: there’s not time to work everything out from scratch

Calculators may be used in the examination: The School of Informatics does not provide calculators for use in exams. If the use of a calculator is permitted in an exam, it’s your responsibility to bring an approved calculator to the exam.

End-of-course feedback:

https://edin.ac/CEQ

Thanks!