Inf2b Learning and Data

Lecture 16: Review

Hiroshi Shimodaira
(Credit: Iain Murray and Steve Renals)

Centre for Speech Technology Research (CSTR)
School of Informatics
University of Edinburgh

http://www.inf.ed.ac.uk/teaching/courses/inf2b/
https://piazza.com/ed.ac.uk/spring2017/infr08009learning
Office hours: Wednesdays at 14:00-15:00 in IF-3.04 → 2.46

Jan-Mar 2017
Today’s Schedule

1. Topic revision
2. Maths formulae to remember
3. Methods/derivations to understand
4. Exam technique
Topics dealt within the course

- Distance and similarity measures (Pearson correlation coef.)
- Clustering (K-means clustering)
- Dimensionality reduction (covariance matrix, PCA)
- Classification
  - $K$-NN classification
  - Naive Bayes
  - Gaussian classifiers (MLE, discriminant functions)
  - Neural networks (Perceptron error correction algorithm, sum-of-squares error cost function, gradient descent, EBP)
- Statistical pattern recognition theories
  - Bayes theorem, and Bayes decision rule
  - Probability distributions and parameter estimation
    - Bernoulli distribution / Multinomial distribution
    - Gaussian distribution
  - Discriminant functions
  - Decision boundaries/regions (minimum error rate classification)
  - Evaluation measures and methods
- Optimisation problems
Maths formulae to remember

- Euclidean distance:
  \[ r_2(x, y) = \| x - y \| = \sqrt{\sum_{i=1}^{D} (x_i - y_i)^2} \]
  
cf. \( \text{sim}(x, y) = \frac{1}{1 + r_2(x, y)} \) as a similarity measure

- Pearson correlation coefficient:
  \[ \rho(x, y) = \frac{1}{N-1} \sum_{n=1}^{N} \frac{(x_n - \mu_x)}{\sigma_x} \frac{(y_n - \mu_y)}{\sigma_y} \]

- Bayes Theorem
  \[ P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \]
  \[ P(C_k|x) = \frac{p(x|C_k)P(C_k)}{p(x)} = \frac{p(x|C_k)P(C_k)}{\sum_{k=1}^{K} p(x|C_k)P(C_k)} \]
Bayes decision rule (cf. MAP decision rule)

\[ k^* = \arg \max_k P(C_k \mid x) = \arg \max_k P(x \mid C_k)P(C_k) \]

Naive Bayes for document classification

(vocabulary: \( V = \{ w_1, \ldots, w_{|V|} \} \), test document: \( D = (o_1, \ldots, o_L) \))

- Likelihood by Bernoulli document model

\[
P(b \mid C_k) = \prod_{t=1}^{|V|} [b_t P(w_t \mid C_k) + (1 - b_t)(1 - P(w_t \mid C_k))]\]

- Likelihood by Multinomial document model

\[
p(x \mid C_k) \propto \prod_{t=1}^{|V|} P(w_t \mid C_k)^{x_t} = \prod_{i=1}^L P(o_i \mid C_k)\]
Maths formulae to remember (cont.)

- **Univariate Gaussian pdf:**
  \[
p(x \mid \mu, \sigma^2) = N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right)
  \]

- **Multivariate Gaussian pdf:**
  \[
p(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right)
  \]

  **Parameter estimation from samples:**
  \[
  \hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n, \quad \hat{\Sigma} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \hat{\mu})(x_n - \hat{\mu})^T
  \]
  NB: \( N \) in case of MLE

- **Correlation coefficient:**
  \[
  \rho(x_i, x_j) = \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} , \quad \Sigma = (\sigma_{ij})
  \]
Maths formulae to remember (cont.)

- Logistic sigmoid function:
  \[ y = g(a) = \frac{1}{1 + \exp(-a)} \]
  \[ g'(a) = g(a)(1 - g(a)) \]

- Softmax activation function (for multiple output nodes):
  \[ y_k = \frac{\exp(a_k)}{\sum_{\ell=1}^{K} \exp(a_\ell)} \]

- 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):
  - Naive Bayes
  - Diagonal covariance matrix rather than a full covariance for each class, shared covariance matrix among classes, regularisation.
  - Dimensionality reduction and feature selection (NE)
Machine learning as optimisation problems

- Euclidean-distance based classification
  \[ k^* = \arg \min_k \|x - r_k\| \]

- K-means clustering
  \[
  \min_{\{z_{kn}\}} \frac{1}{N} \sum_{k=1}^{K} \sum_{n=1}^{N} z_{kn} \|x_n - m_k\|^2
  \]

- Dimensionality reduction to 2D with PCA
  \[
  \max_{u, v} \text{Var}(y) + \text{Var}(z) \\
  \text{subject to } \|u\| = 1, \|v\| = 1, u \perp v
  \]

- 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
  \[ \max_{\mu, \Sigma} L(\mu, \Sigma | D) \]

- Least squares error training of neural networks
  \[
  \min_w \frac{1}{2} \sum_{n=1}^{N} \|y_n - t_n\|^2
  \]
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.
- Solutions are available only for 2008/09, 2009/10, 2013/14 (no plans of releasing those of missing years)
- NB: error in the solution for 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
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 \(\text{ (minus time to pick questions)}\)
- Don’t panic!
- Go for easy marks first
- Don’t spend a long time on any small part
- 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://www.inf.ed.ac.uk/teaching/take-surveys

and

Don’t forget submitting your Inf2b coursework 2 by Wednesday, 05 April 2017 at 4pm

Thanks!