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

Jan-Mar 2014
Today’s Schedule

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

- Distance and similarity measures (collaborative filtering)
- Clustering (K-means clustering)
- Classification
  - K-NN classification
  - Naive Bayes
  - Gaussian classifiers (maximum-likelihood estimation, discriminant functions)
  - Neural networks (Perceptron error correction algorithm, sum-of-squares error cost function, gradient descent, error back propagation)
- 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
  - Evaluation measures and methods
Maths formulae to memorise

- Euclidean distance:
  \[ r_2(x, y) = \| x - y \| = \sqrt{\sum_{d=1}^{D} (x_d - y_d)^2} \]

  cf. \( \text{sim}(x, y) = \frac{1}{1 + r_2(x, y)} \) as a similarity measure

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

- 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)
\[ c^* = \arg \max_{c_k} P(c_k \mid x) = \arg \max_{c_k} P(x \mid c_k)P(c_k) \]

Naive Bayes for document classification
- Likelihood by Bernoulli document model
  \[ P(b \mid c_k) = \prod_{t=1}^{\mid V \mid} [b_t P(w_t \mid c_k) + (1 - b_t)(1 - P(w_t \mid c_k))] \]
- Likelihood by Multinomial document model
  \[ p(x \mid c_k) \propto \prod_{t=1}^{\mid V \mid} P(w_t \mid c_k)^{x_t} \]
Maths formulae to memorise

- 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 by MLE:
  \[ \hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x^{(n)}, \quad \hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} (x^{(n)} - \hat{\mu})(x^{(n)} - \hat{\mu})^T \]

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

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

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

- and basic maths rules (e.g. differentiation)
Methods/derivations to understand (non exhaustive)

- Collaborative filtering
- Clustering and classification
- Discriminant functions of 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.
Exam revision

Look at past papers. There are many.

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

**Recent papers:**
- 2009?

**Don’t overfit!**
Anything that appears in the notes, slides, or tutorial sheets is examinable, unless marked non-examinable, extra topics, or (†)
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
- Know the standard stuff:
  
  there’s not time to work everything out from scratch

(Calculators may be used in the examination)