### Today’s Schedule

1. **Topic revision**
2. **Maths formulae to memorise**
3. **Methods/derivations to understand**
4. **Exam technique**

---

### Maths formulae to memorise

#### Euclidean distance:

\[ r_2(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} \]

- *cf.* \( \text{sim}(x, y) = \frac{r_2(x, y)}{\text{max}(r_2(x, y))} \) as a similarity measure

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

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

- **Bayes decision rule** (cf. MAP decision rule)
  \[ k^* = \arg \max_k P(C_k|x) = \arg \max_k P(x|C_k)p(C_k) \]

  - *Naive Bayes for document classification*
    - Vocabulary: \( V = \{w_1, \ldots, w_N\} \) test document: \( D = \{x_1, \ldots, x_n\} \)
    - Likelihood by Bernoulli document model
      \[ P(b|C_k) = \prod_{i=1}^{n} \left[ b_iP(w_i|C_k) + (1-\beta)(1-P(w_i|C_k)) \right] \]
    - Likelihood by Multinomial document model
      \[ P(x|C_k) \propto \prod_{i=1}^{n} P(w_i|C_k)^{a_i} = \prod_{i=1}^{K} p(a_i|C_k) \]

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

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

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

- **Correlation coefficient:**
  \[ \rho(x_i, y_i) = \frac{\sigma_{x,y}}{\sqrt{\sigma_x \sigma_y}}, \quad \Sigma = (\sigma_{x,y}) \]

#### Logit sigmoid function:

\[ y = g(a) = \frac{1}{1 + \exp(-a)}, \quad g'(a) = g(a)(1-g(a)) \]

- **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)

### 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.

### Machine learning as optimisation problems

- **Euclidean-distance-based classification**
  \[ k^* = \arg \min_k ||x - r_{C_k}|| \]

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

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

- **Least squares error training of neural networks**
  \[ \min_{\theta} \frac{1}{N} \sum_{n=1}^{N} ||y_n - t_n||^2 \]
<table>
<thead>
<tr>
<th>Exam revision</th>
<th>Exam revision (cont.)</th>
<th>Time in the exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look at lecture notes, slides, tutorials, and past papers.</td>
<td>There will be an Inf2b Revision Meeting in April before the exam</td>
<td>• Half an hour per question (minus time to pick questions)</td>
</tr>
<tr>
<td>Early exam papers: many (useful) multiple choice Qs</td>
<td>Date: TBC</td>
<td>• Don’t panic!</td>
</tr>
<tr>
<td>• No longer the exam format</td>
<td>Send me questions/requests that you want me to discuss at the meeting.</td>
<td>• Go for easy marks first</td>
</tr>
<tr>
<td>• Syllabus has changed slightly</td>
<td></td>
<td>• Don’t spend a long time on any small part</td>
</tr>
<tr>
<td>Recent exam papers since 2008/09</td>
<td></td>
<td>• Know the standard stuff:</td>
</tr>
<tr>
<td>• Solutions are available only for 2008/09, 2009/10, 2013/14 (no plans to release those of missing years)</td>
<td>there’s not time to work everything out from scratch</td>
<td>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.</td>
</tr>
<tr>
<td>• NB: error in the solution for 5 (c) of 2008/09: square root is not taken in computing standard deviations.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Don’t overfit!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anything that appears in the notes, slides, or tutorial sheets is examinable, unless marked non-examinable, extra topics, or (†)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Don’t trust unofficial solutions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inf2b Learning and Data: Lecture 16 Review 11