COMS21202 Symbols, Patterns and Signals

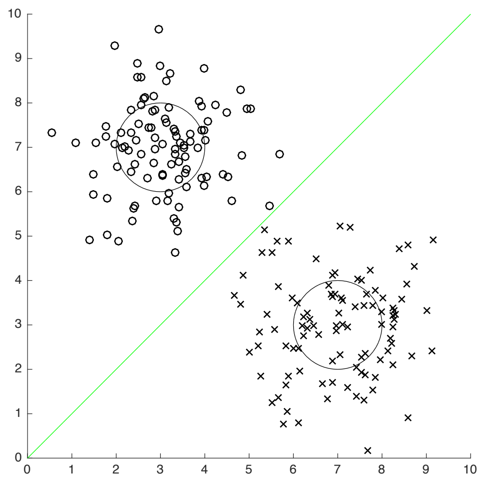
Progress test 2

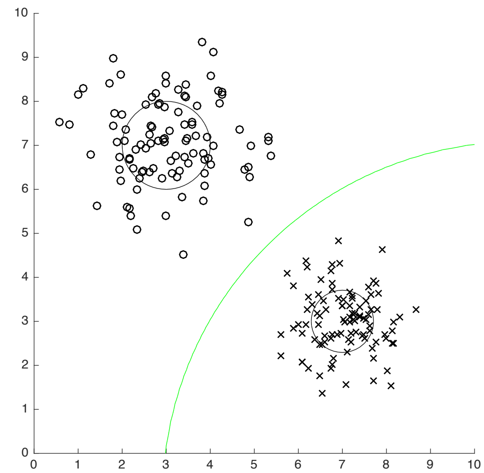
This test is worth 5% of the unit mark, and is marked out of 20.
Answer the following questions in the boxes on the sheet provided (do not hand in additional sheets).
The data given varies per sheet.

For each question, **show how you arrived at your answer**.

1. Each of the two diagrams on the answer sheet shows a two-class classification problem, with estimated means and covariances indicated by ellipses. In each diagram, sketch the decision boundaries if we use these estimates for maximum-likelihood classification. Provide brief explanations. *[5 marks]*

   Answer:





   Explanation: (left) When both covariances are equal the decision boundary is linear, and when there is no correlation between the features it is the perpendicular bisector. (right) Different covariances so non-linear decision boundary, curving towards the tighter distribution.

2. The answer sheet gives a data set consisting of six instances described by four Boolean features, and labelled with classes $+$ or $-$. We want to learn a decision tree using the four features A, F, Q and X.

   (a) Before doing any calculations, what can you say about the relative quality of these features – which ones do you expect to be better, worse or equally good for splitting on at the root of the tree? Justify your answer in words. *[3 marks]*

   (b) Information gain is defined as the difference between the impurity of the parent and the weighted average impurity of the children, where impurity is measured by entropy. Calculate information gain for A and X. Show your working. *[4 marks]*

   Answer:

   (a) A and F are equally good by symmetry. Q and X are equally good by symmetry. A, F are better because one child is pure.

   (b) For A the calculation is as follows:

$$1-\frac{4}{4+2}\left(-0.75\log_2 0.75-(1-0.75)\log_2(1-0.75)\right)-\frac{2}{4+2}(0)=0.46$$

For X it is:
$1-\frac{3}{3+3}(-0.67\log_2 0.67-(1-0.67)\log_2(1-0.67))-\frac{3}{3+3}(-0.33\log_2 0.33-(1-0.33)\log_2(1-0.33))=$
0.082.

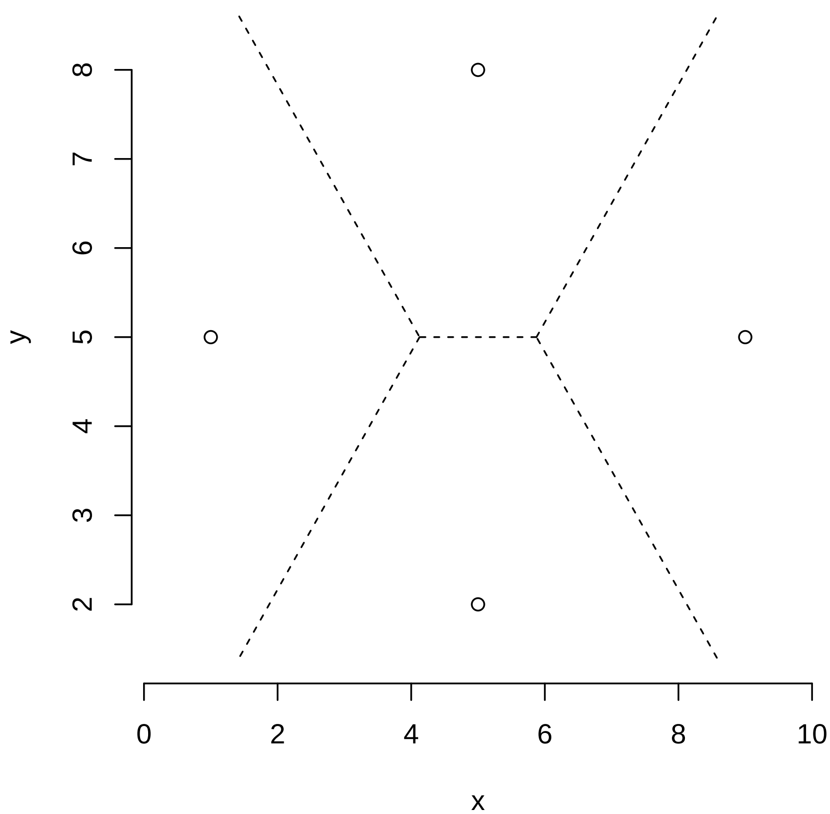
3. The answer sheet gives four 2-D data points **r**, **s** **t** and **u**.

   (a) Plot these points and sketch the 1-nearest neighbour decision boundaries (Voronoi tesselation). *[2 marks]*

   (b) Also given is a query point **q**. If **r** and **s** are labelled positive and **t** and **u** are labelled negative, classify **q** using
   (i) 1-nearest neighbour classification, and
   (ii) 3-nearest neighbour classification, both with Euclidean distance. *[4 marks]*

   (c) How would your answers be affected if you used Manhattan distance instead? *[2 marks]*

Answer:



(a)

(b) **q** *is closest to* **s** *(distance* 2*), then* **t** *and* **r** *(distance* 3.61*), then* **u** *(distance* 6*). So 1nn will classify it as positive and 3nn as negative.*

(c) The only distances that change are with **t** *and* **u** *which are now at distance* 5*, but this doesn't change the classifications.*

End of questions. Formula sheet overleaf

COMS21202 Symbols, Patterns and Signals

Progress test 2: Formula sheet

Given a covariance matrix Σ, the Mahalanobis distance between vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathrm{Dis}_M(\mathbf{x},\mathbf{y}|\Sigma) = \sqrt{(\mathbf{x}-\mathbf{y})^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mathbf{y})}$$

The multivariate normal distribution can be expressed in terms of Mahalanobis distance as follows:

$$P(\mathbf{x}|\mu,\Sigma) = \frac{1}{E_d}\exp\left(-\frac{1}{2}\left(\mathrm{Dis}_M(\mathbf{x},\mu|\Sigma)\right)^2\right), \quad E_d = (2\pi)^{d/2}\sqrt{|\Sigma|}$$

Probabilistic classification

Bayes' rule: $P(\omega|\mathbf{x}) = \dfrac{P(\mathbf{x}|\omega)P(\omega)}{P(\mathbf{x})}$

Maximum a posteriori decision rule: $\omega_{\mathrm{MAP}} = \arg\max_{\omega} P(\omega|\mathbf{x}) = \arg\max_{\omega} P(\mathbf{x}|\omega)P(\omega)$

Maximum likelihood decision rule: $\omega_{\mathrm{ML}} = \arg\max_{\omega} P(\mathbf{x}|\omega)$

Naive Bayes decision rule (ML): $\omega_{\mathrm{nB}} = \arg\max_{\omega} \prod_{i=1}^{d} P(x_i|\omega)$

Distances

Minkowski metric: $L_k(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{d}|x_i - y_i|^k\right)^{1/k}$

Special cases: Manhattan distance $L_1(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{d}|x_i - y_i|$; Euclidean distance $L_1(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2}$; Chebyshev distance $L_\infty(\mathbf{x},\mathbf{y}) = \max_{i=1}^{d}|x_i - y_i|$.

Decision trees

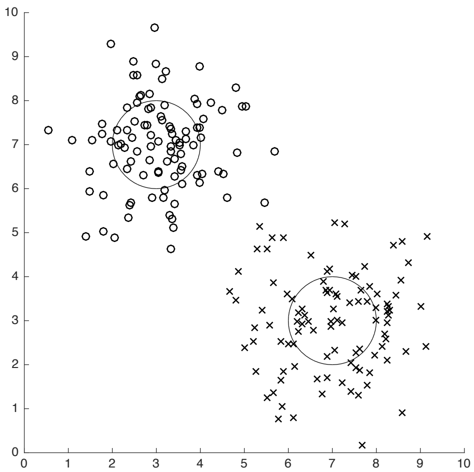
Entropy (c classes, proportion of j-th class is p_j): $\sum_{j=1}^{c} -p_j \log_2 p_j$

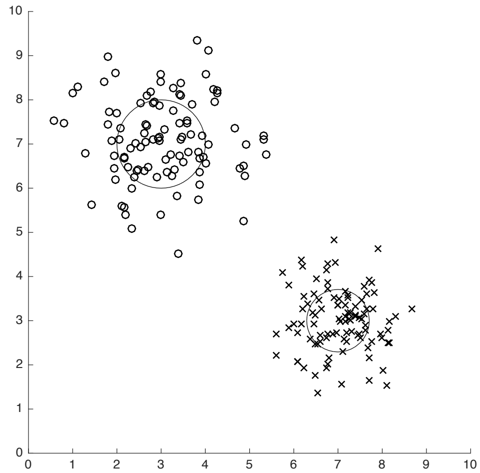
Information gain: $\mathrm{entropy}(\mathit{Parent}) - \sum_{i=1}^{k}\frac{n_i}{N}\,\mathrm{entropy}(\mathit{Child}_i)$

COMS21202 Symbols, Patterns and Signals

Progress test 2: Answer sheet

Name: .... Username: .... Course: ....





Q1.

Q2.

A	F	Q	X	Class
0	0	1	1	+
0	1	1	0	+
0	0	0	0	+
0	1	0	0	–
1	1	1	1	–
1	1	0	1	–

Please give your answer to Q3 overleaf

Q3.(a) $\mathbf{r} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$, $\mathbf{s} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$, $\mathbf{t} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$.

Q3.(b-c) $\mathbf{q} = \begin{bmatrix} 7 \\ 5 \end{bmatrix}$