EGT2
ENGINEERING TRIPOS PART IIA

Wednesday 26 April 2017 2 to 3.30

Module 3F8

INFERENCE

Answer not more than ***three*** *questions.*

All questions carry the same number of marks.

The ***approximate*** *percentage of marks allocated to each part of a question is indicated in the right margin.*

Write your candidate number ***not*** *your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

1 (a) Explain what is *maximum likelihood estimation* and how it is used to estimate parameters in a probabilistic model from data. [20%]

(b) A source emits N signals x_n drawn independently from a Gaussian distribution with mean 1 and variance 1. The signals are measured by a receiver a fixed distance d metres away. The signals are exponentially attenuated and corrupted by independent Gaussian noise so that the measurements are given by

$$y_n = \exp(-d)\, x_n + \varepsilon_n.$$

The noise ε_n has zero mean and variance 1.

The formula for a one dimensional Gaussian distribution with mean μ and variance σ^2 is

$$p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right).$$

(i) Compute the mean and the variance of a single measurement y_n under the probabilistic model. [30%]

(ii) Use your answer to (i) to compute the likelihood of the parameter d when N measurements have been made. [20%]

(iii) Four measurements are made $\{y_n\}_{n=1}^{N} = \{1, -2, -1, 2\}$. Find the maximum likelihood setting of the parameter d for these data. You may find it simpler to first find the maximum likelihood setting for $w = \exp(-d)$ and then rearrange to find the estimate of d. [30%]

2 (a) Compare and contrast *regression* and *classification* tasks in machine learning. [30%]

(b) A dataset comprises pairs of real valued inputs x_n and real valued outputs y_n shown in Fig. 1. Suggest a suitable probabilistic model for these data that could be used to predict an output from a new input. Explain your reasoning. [30%]

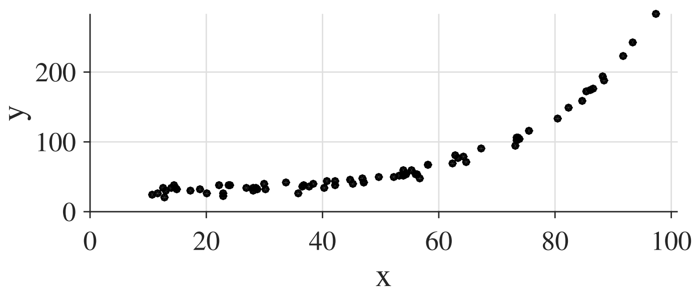

Fig. 1

(c) A second set of discrete valued outputs z_n, shown in Fig. 2, were measured simultaneously with y_n so that the training data is now $\{x_n, y_n, z_n\}_{n=1}^{N}$. Extend the probabilistic model you proposed for part (b) so that it can be used to jointly predict both outputs from a new input. Explain your reasoning. [30%]

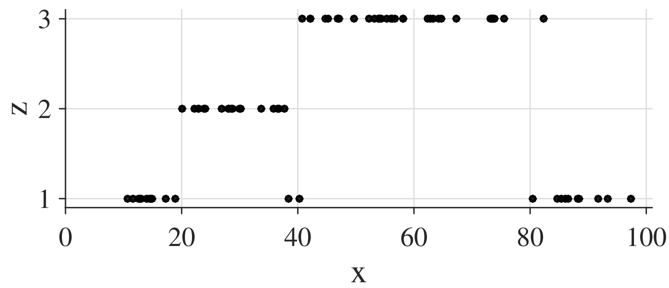

Fig. 2

(d) Consider the extended model you have proposed in part (c). Do the second set of outputs provide useful information about the parameters of the original component described in part (b)? Explain your reasoning. [10%]

3 (a) Describe what *clustering* is and give an example application where a clustering algorithm might be used. [20%]

(b) A simple one dimensional dataset, $\{y_n\}_{n=1}^{N} = \{-10.1, -9.9, 9.9, 10.1\}$, is modelled using a mixture of Gaussians. The mixture comprises two components with class membership probabilities $p(s_n = 1) = \alpha$ and $p(s_n = 2) = 1 - \alpha$. The component distributions are given by $p(y_n|s_n = 1, \theta) = \mathcal{N}(y_n; \mu_1, \sigma_1^2)$ and $p(y_n|s_n = 2, \theta) = \mathcal{N}(y_n; \mu_2, \sigma_2^2)$ where

$$\mathcal{N}(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right).$$

The parameters of the model are collectively denoted $\theta = \{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \alpha\}$.

(i) The model is fitted using the Expectation Maximisation (EM) algorithm. The posterior distribution over the class labels for the nth data point is denoted $r_{n,k} = p(s_n = k|y_n)$. Write down the algorithm's M-Step update equations. [40%]

(ii) The EM algorithm is run to convergence, returning a parameter estimate θ_{EM}. The estimate is found to depend on the initialisation. Sketch the various Gaussian mixture model fits that the EM algorithm returns, i.e. sketch the densities $p(y|\theta_{\text{EM}})$ as a function of y. Where possible, indicate the estimated parameter values approximately. Identify which estimates correspond to global optima of the likelihood. [40%]

4 (a) Explain what the terms *Markov property*, *filtering* and *stationary distribution* refer to in the context of *Hidden Markov Models*? [30%]

(b) A probabilistic model for a time-series containing binary valued observations y_n employs binary state variables s_n. The transition matrix and emission matrix of the model are denoted

$$T = \begin{bmatrix} p(s_{n+1}=0|s_n=0) & p(s_{n+1}=0|s_n=1) \\ p(s_{n+1}=1|s_n=0) & p(s_{n+1}=1|s_n=1) \end{bmatrix}, \; E = \begin{bmatrix} p(y_n=0|s_n=0) & p(y_n=0|s_n=1) \\ p(y_n=1|s_n=0) & p(y_n=1|s_n=1) \end{bmatrix}.$$

The forward filtering recursions have been used to process N observations, $y_{1:N}$, in order to return the posterior distribution over the Nth state variable,

$$\rho = \begin{bmatrix} p(s_N=0|y_{1:N}) \\ p(s_N=1|y_{1:N}) \end{bmatrix}.$$

(i) Explain how to transform the posterior distribution over the Nth state into a forecast for the observations one time step into the future, i.e. express $p(y_{N+1}|y_{1:N})$ in terms of ρ. [25%]

(ii) Now provide a forecast for the observations τ time steps into the future by expressing $p(y_{N+\tau}|y_{1:N})$ in terms of ρ. [15%]

(iii) Compute the forecast $p(y_{N+\tau}|y_{1:N})$ in the limit $\tau \to \infty$ when

$$T = \begin{bmatrix} 3/4 & 1/2 \\ 1/4 & 1/2 \end{bmatrix}, \; E = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}.$$

Explain your reasoning. [30%]

END OF PAPER

THIS PAGE IS BLANK

Analytic answers:

1b
i) mean$(y) = \exp(-d)$, variance$(y) = 1 + \exp(-2d)$
ii) $\log p(y_{1:N}|d) = -\frac{N}{2} log(2\pi(1+\exp(-2d))) - \frac{1}{1+\exp(-2d)} \sum_n (y_n - \exp(-d))^2$
iii) $d = \frac{1}{2}\log(3) = 0.347$ (3 s.f.)

3b
i) $\pi_k = \frac{1}{N}\sum_n r_{nk}$, $\mu_k = \frac{1}{\sum_n r_{nk}} \sum_n r_{nk} y_n$, $\sigma_k^2 = \frac{1}{\sum_n r_{nk}} \sum_n r_{nk}(y_n - \mu_k)^2$

4b
i) $p(y_{N+1}|y_{1:N}) = ET\rho$
ii) $p(y_{N+1}|y_{1:N}) = ET^{\tau}\rho$
iii) $p(y_\infty) = \frac{1}{12}\begin{bmatrix} 7 \\ 5 \end{bmatrix}$