Q1:

a) Some emails received by users are spams, attempting to gain illicit access to employee’s accounts. You are building a classifier to detect such illicit emails. You start by using the feature of whether or not an email asks for the recipients password, as this an important datum indicating if the email is in fact a hacking attempt. In particular 95% of hacking emails contain requests for a user’s password, and 95% of legitimate emails do not contain requests for a user’s password. 1% of emails overall are hacking attempts.

i. If your classifier scans an email with a password request, what is the probability that the email is in fact a hacking attempt?

ii. Comment on this value that you calculate.

iii. What is the probability that your classifier makes an error? [7+2+3=12 marks]

b) Explain the difference between Maximum Likelihood (ML) and Maximum a Posterior (MAP) methods of parameter estimation. [4 marks]

c) Give the equations for ML and MAP estimates for the bias of a coin. [3 marks]

d) Explain the underlying probabilistic assumptions that implicitly underlie (i) the least squares error function and (ii) weight regularization in linear regression. [6 marks]

Q2:

a) Complete this table for the methods of linear regression and logistic regression. Give (on your answer sheet) the form of the model to be learned, the cost function, and a possible method (gradient or exact) for minimising this cost function. [12 marks]

|  |  |  |  |
| --- | --- | --- | --- |
|  | Form of model | Cost function | Optimisation strategy |
| Linear Regression |  |  |  |
| Logistic Regression |  |  |  |

b) Consider a three-layer neural network with d input units, h hidden units, c output units, and bias. The network has activation function f , with input-to-hidden weights labelled w\_ji and hidden-to-output weights labelled w\_kj . For a single input vector x = (x1,x2,...,xd) with associated target vector t = (t1,t2,...,tc), write down an expression for:

i. the output activation yj of a hidden unit;

ii. the output activation zk of an output unit (i.e. an output of the network);

iii. the squared error J. [3+3+3=9 marks]

c) Practical challenges with training neural networks include: (i) getting stuck in local optima, (ii) underfitting or overfitting. Explain the practical strategies you would use to overcome these issues. [4 marks]

Q3:

(a) The probability density function of the Gaussian distribution is given by:  
  
p(x) = [(1) / (2pi^{m/2} \* E^{0.5})] \* exp^(-0.5\*(x-u)^T \* inverse sum of (x-u))  
  
Explain what each symbol represents and give its dimensionality. Sketch the probability density function in the 1 dimensional case and use it to support your explanations. [6 marks]

(b) The maximum likelihood principle for density estimation states that the optimal parameters 𝜃 ∗ are the ones that minimise the following quantity:  
  
E = - sum of (inverse log (p(x\_n; theta)))

i. Explain the terms in the equation above, and state what are the parameters theta in the case of a Gaussian.

ii. Explain how the derivatives of this function can be used to optimise 𝐸 in the equation above and show that in the case of the mean vector u this leads to:  
  
u = 1/N \* sum of (x\_n)

For simplicity you may do the derivations for the 1-dimensional case, that is, in the case that the data are scalars and the covariance matrix a scalar as well. [4+4 = 8 marks]

(c) Explain how density estimation with a single Gaussian can be used in a practical application. Pick one of the two examples below or devise your own problem. Explain the steps carefully. For the example that you choose explain whether you could have used a Supervised Learning algorithm and, if not, why.

i. Background modelling and subtraction in a Computer Vision application where the goal is to determine at each frame which pixels belong to the foreground and which to the background.

ii. Abnormality detection in any context. For example in the context of consumer buying habits in a supermarket chain. [4+4=8 marks]

(d) In several cases the covariance matrix is singular and cannot be inverted. Why is this a problem? Give an example, perhaps with the use of a diagram, and describe a solution to that problem. [3 marks]

Q4:

a) With a help of a diagram explain the main principles of the first-order Markov Model. Explain what is meant by the term ‘’first-order”. What are the differences with a hidden Markov model (HMM)? In your answer, define the states w\_i , the symbols v\_k , and the matrices A = [a\_ij] and B = [b\_jk] . [6 marks]

b) The decoding problem can be stated as follows: Given an HMM and a sequence of observation symbols V^{1:T} determine the most likely sequence of hidden states w^{1:T} .What are the other two types of problems considered in the context of HMMs? [6 marks]

(c) You are given the task to design a speech recogniser that distinguishes between the words “yes” and “no”.

i. Describe how the k-means algorithm can be used to transform a speech signal to a sequence of symbols {𝑣1, 𝑣2, … , 𝑣𝐾}. (Hint: As a first step divide the signal in small chunks of equal and fixed length)

ii. Describe how you would design the training process. Make reference to what kind of data you will have, which algorithm you will use (e.g. refer to your answer to part (b)) and how many HMMs you need to train. Why can’t one use a simple classifier, such as a linear logistic regression scheme?

iii. Describe how you would use the HMMs that you trained in order to make decisions at test time. [4+4+5 = 13 marks]