Q1:

a) Define the joint probability P(A,B) and the conditional probability P (A given B). You may want to use a diagram/sketch. Give the formula that relates them. [3 marks]

b) Give the relation between the joint probability P (X , Y) and the probabilities P (X ) and P (Y ) that holds in the case that X and Y are independent random variables. Give the condition that holds when X and Y are uncorrelated. Are these conditions the same? [6 Marks]

d) An IT worker works from home 2 days a week. When she works from home there is a 30% chance she will not answer an email within an hour, 10% chance that she will not answer an email within two hours, and it is certain that she will answer all the emails within the day. When she is at office, there is a 50% chance she will not answer an email within one hour, 10% chance that she will not answer an email within the two hours, and it is certain that she will answer all the emails within the day.

i. If you send her an email, what is the probability that she will answer within 2 hours? Justify your answer.

ii. Given that she hasn’t replied to your email within 1 hour, what is the probability that she is working from home? Does the information that she hasn’t answered the email within 1 hour makes it more or less likely that she works from home? [12 Marks]

d) Explain the difference between Maximum Likelihood (ML) and Maximum a Posterior (MAP) methods of parameter estimation. [6 marks]

Q2:

a) Compare and contrast the goals in Linear Regression and Logistic Regression. [4 marks]

b) The form of a linear regression model is y=w^T\*x. Assuming the mean squared error cost function, derive gradient descent updates for the weights w. [9 marks]

c) What is the limitation of the networks without hidden layers, that was overcome by Multilayer Networks? Is it is essential that the activation function is non-linear? [6 marks]

d) Practical pitfalls 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. [6 marks]Q3:

(a) Describe the difference between supervised and unsupervised learning. Give an example of a real world problem that requires a supervised learning algorithm and an example of a real world problem that can be solved with an unsupervised learning algorithm. In both cases define the inputs and the outputs. [8 marks]

(b) Describe in detail the steps of the K-means algorithm. Make sure that you define the input to the algorithm, the output, and the dimensionality of all the variables that you use. [8 marks]

(c) Identify the two sets of variables that are estimated by the K-means algorithm. Explain what coordinate descent (or coordinate optimisation) is. Using a sketch, show that this general optimisation method is warranted to converge. [4 marks]

(d) The K-Means algorithm converges to a local minimum. Describe a practical method to deal with this problem. Can this method be used to determine the optimal value of K? [5 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]