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]

c) You are building a system to detect fraudulent card transactions. The first feature that you use is whether the transaction has taken place in the country where the card has been registered (country of registry). Data analysis suggests that 95% of fraudulent transactions do not take place in the country of registry, 98% of the genuine transactions take place in the country of registry, and 1% of transactions overall take place in a country different to the country of registry.

i. What is the probability that a transaction is taking place in a country different than the country of registry?

ii. If a transaction has not taken place in the country of registry, what is the probability that it is fraudulent?

iii. What is the decision that your classifier makes for a transaction took place in country different than the country of registry? What is the probability that the classifier makes an error for such a transaction? Is this the only type of error that the classifier makes? Explain your answer. [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) The form of non-linear regression is 𝑦 = 𝒘^T\*phi (𝒙), where phi(.) are non-linear basis functions.

i. Give two examples of basis functions. Give an example of a basis function that would be appropriate for modelling the temperature at Hyde Park.

ii. What learning algorithms can be used to learn the model parameters in the case of non-linear regression?

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) An advertising company that wants to do market segmentation, decides to do so using the K-means algorithm. A colleague suggests that in order to determine the optimal number of segments/clusters, the company should run the K-Means algorithm several times for different values of K and keep the value of K that gives the lowest value of the objective function. Is she/he right and why? Give an alternative method for determining 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. [6 marks]

d) What is the evaluation problem? Using the results of (c), or otherwise, present a naïve algorithm that solves the evaluation problem. Can this algorithm be used in practice? (Hint: Use the law of total probabilities) [7 marks]