Q1:

a) A Machine Learning practitioner needs to estimate the parameters theta of a model, given some data 𝑋. Suppose, that the parameters can take only two values, theta 1 or theta 2 and that it is given that 𝑝(𝑋 given theta 1) = 0.6, 𝑝(𝑋 given theta2) = 0.5, 𝑝(theta 1) = 0.3, and 𝑝(theta 2) = 0.7.

i) What does each of the terms 𝑝(𝑋 given theta𝑖) and 𝑝(theta 𝑖), 𝑖 contained in {1,2} denote?

ii) Which model will he/she choose under the Maximum Likelihood criterion? Which model will he/she choose under the Maximum Aposteriori Probability criterion? Show the calculations. [8 marks]

b) Explain the “gradient descent” algorithm. What is the learning rate in this context and how does a Machine Learning practitioner set its value? [9 marks]

c) In a Machine Learning problem, you are asked to fit a linear regression model to data from a dataset 𝑋 = {𝑥1, 𝑥2, … , 𝑥𝑚}, where 𝑥𝑚 ∈ 𝑅 𝑑 , with 𝑚 = 10^5 and 𝑑 = 10^8 . i) What is the specific issue that arises given those values of 𝑚 and 𝑑 and how can you address it? ii) Give the equation that you will use to estimate the parameters of the model in that case and explain what each term denotes. [8 marks]

Q2:

(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. [8 marks]

b) Fit a Gaussian distribution to the following set of 1-dimensional data 𝑋 = {3,2,4,4,2,3}. Show the formulas that you use for your estimation and state what is the criterion that is optimised. [10 marks]

c) In several cases the covariance matrix is singular and cannot be inverted. In which cases is this more likely to happen? Why is this a problem? Give a solution to it. [7 marks]Q3:

a) You need to design a Neural Network that solves the problem of facial attribute recognition. More specifically the network should receive in the input an image of a face, and should recognise whether the depicted subject wears glasses or not, has long or short hair, smiles or not and should recognise its apparent age. Design the first and the last layers of such a network, detailing your choices. Define the total cost function and give the format of a training example and the corresponding ground truth associated with it. [Hint: You can treat the recognition of the age either as a regression problem, or as a classification problem – either choice is equally valid.] [12 marks]

b) How can one decide whether the network is overfitting or underfitting the data? Describe the process that one needs to follow. [6 marks]

c) Why are the neural networks weights initialised with random values? [7 marks]

Q4:

a) Describe the difference between supervised and unsupervised learning. Give an example of a supervised learning problem and an example of an unsupervised learning problem. [6 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. [10 marks]

c) After applying K-Means clustering on unlabelled data, a Machine Learning practitioner runs a linear classifier using as pseudolabels the cluster assignments. Why would he/she want to do that? What type of regularisation would you advice her/him to use and why? [9 marks]