IMPERIAL COLLEGE LONDON

MEng and MSc EXAMINATIONS 2021

Part IV and Advanced Mechanical Engineering

for Internal Students of the Imperial College of Science, Technology and Medicine
This paper is also taken for the relevant examination for the Associateship or Diploma

MACHINE LEARNING

Monday, 26th April; Duration: 90 minutes

This paper contains SIX questions.
Attempt every question.
The numbers shown by each question are for your guidance; they indicate approximately how the examiners intend to distribute the marks for this paper.
A Data and Formulæ Book is provided.

This is an OPEN BOOK Examination.

This time-limited remote assessment has been designed to be open book. You may use any resources which you have available. You may use any tools including, but not limited to, graphical calculators, spreadsheets and programming software. Mathematical derivations, if required, should show all assumptions made and contain detailed explanations of each step. Calculations must clearly show any formulae used and all intermediate steps.

The use of the work of another student, past or present, constitutes plagiarism. Giving your work to another student to use constitutes an offence. Collusion is a form of plagiarism and will be treated in a similar manner. This is an individual assessment and thus should be completed solely by you. The College will investigate all instances where an examination or assessment offence is reported or suspected, using plagiarism software, vivas and other tools, and apply appropriate penalties to students. In all examinations we will analyse exam performance against previous performance and against data from previous years and use an evidence-based approach to maintain a fair and robust examination. As with all exams, the best strategy is to read the question carefully and answer as fully as possible, taking account of the time and number of marks available.

Turn over

1. My company has been making carbon fibre components and I have found that 1 in 10 are breaking. I am aware that there are variations in our production process, and I wish to use information from this to predict failure. In particular I wish to use the temperature at which the carbon fibre has been cured. Below is a representative sample of the curing temperatures for components which did not break:

252.9 212.0 229.4 267.2 256.0 170.7 228.5 195.5 196.9 212.3 204.3 243.6
222.8 203.7 213.3 210.0 244.8 193.8 209.4 174.4 123.4 219.6 225.9 177.7
268.1 156.4 201.4 194.4 246.0 244.1 204.6 211.3 173.4 140.6 189.6 204.7
236.9 236.1 188.4 190.9 168.5 157.4 148.8 258.5 184.7 186.9 162.4 223.3
151.6 193.6 173.1 211.6 184.7 164.6 199.2 212.8 202.0 209.1 181.0 189.1
179.8 189.2 175.6 148.2 205.3 187.9 151.1 213.9 172.8 201.6 221.9 203.9
234.2 163.0 212.1 179.5 173.9 182.6 190.7 201.7 165.0 227.0 214.0 153.9
244.6 256.9 235.4 194.6 167.9 231.6 187.9 236.7 206.2 229.3 210.7 221.2
200.3 253.6 203.8 212.1

The following values are derived from this dataset and you may find these useful:

- Sum: 20179.5
- Sum of squares: 4163531.8
- Median: 202.8
- Number of points: 100

Below is a representative sample of the curing temperatures of a sample of components which broke:

334.7 189.4 192.8 293.6 197.2 337.5 231.4 216.4 336.5 316.6 334.0 290.8
211.2 336.0 237.9 286.1 292.6 243.0 277.6 291.5 266.9 200.5 263.4 309.7
218.7 243.3 230.4 333.2 280.3 268.3 215.4 274.3 219.7 251.4 221.4 280.4
275.9 240.6 267.8 200.8 182.9 269.8 257.5 278.6 357.2 292.5 208.9 300.3
190.8 229.2 246.9 327.1 216.5 212.8 245.6 220.1 300.7 201.4 198.4 230.3
227.6 336.8 292.7 253.9 194.9 288.0 205.0 180.5 303.5 264.3 291.4 264.3
288.6 220.7 203.5 280.7 213.8 219.0 229.5 250.8 234.1 188.1 221.0 149.9
278.1 177.9 200.3 252.3 216.7 319.4 191.8 262.0 248.2 197.4 273.5 242.3
284.7 287.1 347.3 310.1

The following values are derived from this dataset and you may find these useful:

- Sum: 25368.4
- Sum of squares: 6652312.5
- Median: 251.1
- Number of points: 100

All values are in °C (and $(°C)^2$ for the sum of squares).

Assuming that both of the probability distributions for breaking and non-breaking cases can be well captured by a normal distribution, if I take a component where the curing temperature was 240 °C, estimate the probability that the component will break. [15%]

2. You are given the following points: (1, 1), (2, 2), (2, 3).

   (a) Treat this as a supervised learning problem and perform a linear regression on these points, outputting the equation of the best-fit line. [5%]

   (b) Now treating it as an unsupervised problem, perform Principal Component Analysis (PCA) and write out the equation of the line passing through the data in the direction of the first principal component. [10%]

   (c) Explain, using a diagram where appropriate, why the gradient of the line is different in the two cases. Also explain how this relates to regression being a supervised method and PCA being unsupervised. [15%]

3. Sketch out a decision tree corresponding to the decision function in Figure Q3, where white corresponds to class A and black to class B, making clear at each node what the equations for the corresponding inequalities are. To receive full marks your decision tree should be as concise as possible. [10%]

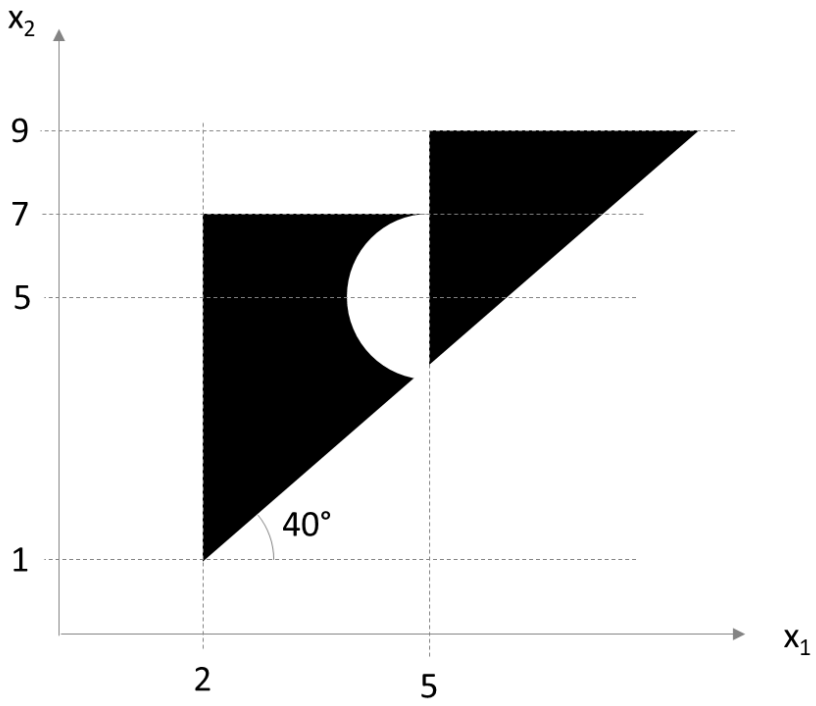

Figure Q3

4. I have points A and B at (0, 0) and (3, 2) respectively.

   (a) Find the coordinates (Cx, Cy) of the point C which is equidistant from both A and B, as measured under both the L1 and L∞ norm, subject to the limits $0 < Cx < 2$ and $2 < Cy < 4$. [20%]

   (b) Show that the L2 distances are also equal at this point. [5%]

5. I wish to undertake a k-means clustering approach. At one iteration I have means at (3, 5) and (7, 2).

   (a) Sketch out a single node neural network which would classify points as -1 or +1 depending on which respective mean they should be allocated to at this iteration, -1 indicating the point being allocated to (3, 5) and +1 for allocation to (7, 2). Calculate all the weighting terms in the network. Make sure you clearly state any activation functions used. [12%]

   (b) Explain why additional layers and nodes are not needed for this problem. [4%]

6. I have two principal components of a system, which are (2/3, 1/3, 2/3) and (1/3, 2/3, -2/3). List all other principal components. [4%]