Calculators may be used in this examination provided they are not capable of being used to store alphabetical information other than hexadecimal numbers

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Visualisation

Main Summer Examinations 2023

Time allowed: 2 hours

[Answer all questions]

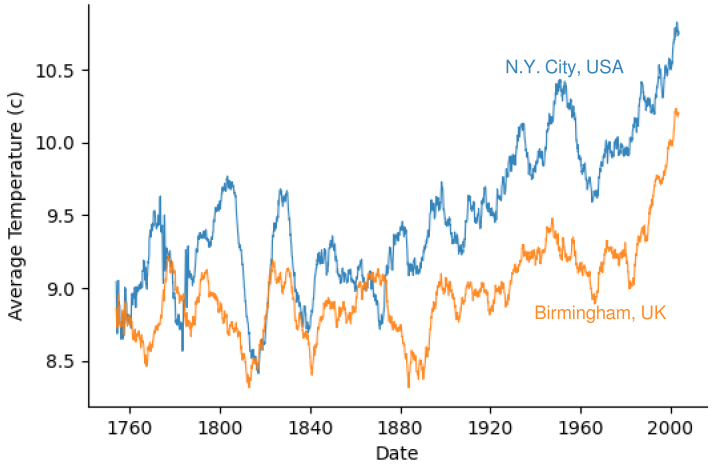
Turn Over

Note

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 60, which will be rescaled to a mark out of 100.

Question 1

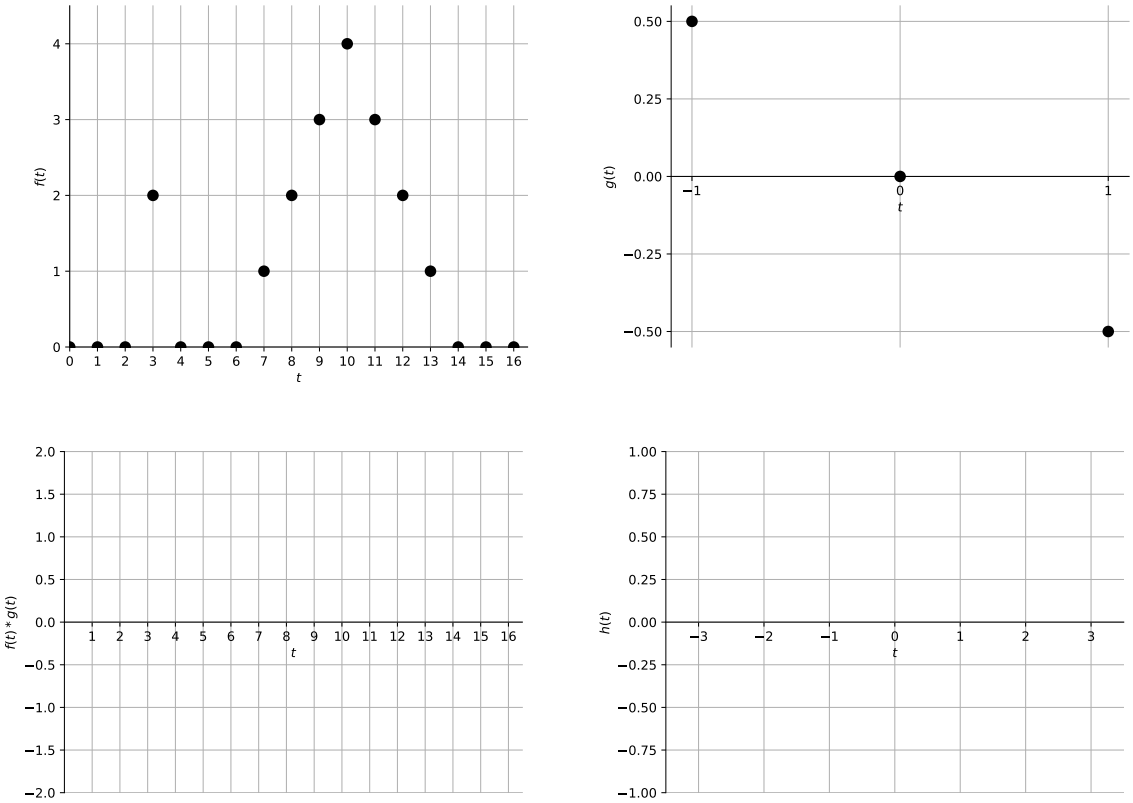
(a) Have a look at the plot below. Three different Gestalt principles have been used in the design of the plot. Name them and explain how they have been used. Keep your answer brief (no more then 1 sentence per Gestalt principle). **[9 marks]**



(b) Please list, which variables have been mapped to which aesthetics using which type of scale. **[3 marks]**

(c) What could be a story you could tell with this data and how could you modify the chart to tell a story, as suggested by Nussbaumer Knaflic? Keep you answer brief (3 to 4 bullet points describing the changes). **[8 marks]**

Question 2



(a) In the top part of the figure above, you can see two functions $f(t)$ and $g(t)$. Please copy the empty chart on the bottom left into your answer book. You don't have to draw all the grid lines. Convolve the two functions and draw the results into your chart.

[9 marks]

(b) The values in function $g(t)$ are chosen such that convolving any function with $g(t)$ computes an approximation of the function's first derivative. A simple way to compute the second derivative is to convolve the result again with function $g(t)$, thus computing the derivative of the derivative. However, this requires two convolution operations. It is possible to construct a function $h(t)$, such that a convolution with $h(t)$ is identical to convolving twice with function $g(t)$. How can you construct such a function? Please answer in two sentences or less. Please copy the empty chart on the bottom right of the above figure into your answer book. You don't have to draw all the grid lines. Draw the function $h(t)$ into your empty chart. Tip: you can extend $g(t)$ with zeros in both directions. **[5 marks]**

(c) A data scientist wants to apply $g(t)$ to their time series to compute the derivative. However, they find that their data is too noisy and requires smoothing. They decide

to use a frequency based low pass filter, by transforming the data into the frequency domain, multiplying filter weights to the coefficients and then applying the inverse Fourier transform. Should they first convolve with $g(t)$ and then apply the low pass filter, or first apply the low pass filter and then convolve with $g(t)$? Briefly explain your reasoning (max 3 sentences). **[3 marks]**

(d) Can you think of a way to apply the smoothing and compute the derivative at the same time using only single operation? Briefly explain your reasoning (max 2 sentences). **[3 marks]**

Question 3

We would like to use direct principal component analysis (PCA) or dual PCA to reduce the dimension of n, d-dimensional dataset $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$. If we define

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 \\ -1 & 1 & 1 & 2 & 2 \end{bmatrix}.$$

(a) For direct PCA, what would be the co-variance matrix $\mathbf{S}$ of $\mathbf{X}$? Please show your derivation to compute this matrix numerically (including any intermediate matrices you calculate). **[4 marks]**

(b) The eigenvalues and eigenvectors of $\mathbf{S}$ are given as follows:

$$\lambda_1 = 10, \mathbf{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \lambda_2 = 2, \mathbf{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Assume we use direct PCA to project $\mathbf{X}$ into $\mathbf{Y}$ such that the variance of the projected dataset $\mathbf{Y}$ has the minimum variance. Please work out in this case how to compute $\mathbf{Y}$ numerically. Please present your results using the fraction style (not decimal style). **[4 marks]**

(c) Now, please show your derivation regarding how to recover the dataset $\tilde{\mathbf{X}}$ from $\mathbf{Y}$ such that it has the same dimension of the original $\mathbf{X}$. Next, please explain (1) why $\tilde{\mathbf{X}}$ is not strictly equal to $\mathbf{X}$ and (2) how we can recover the exact $\mathbf{X}$ from $\mathbf{Y}$? Please present your results using the fraction style (not decimal style). **[4 marks]**

(d) In dual PCA, we need to compute a similar square matrix as $\mathbf{S}$. What would be this matrix? Please show your derivation to compute this matrix numerically. Next, please prove that in terms of training data projection how dual PCA is derived from direct PCA using the concept of singular value decomposition. **[4 marks]**

(e) Now we want to use direct PCA to reduce the dimension of a new dataset of 10 dimensions (i.e., $d = 10$). Assume that the first five eigenvalues of the co-variance matrix are respectively $\lambda_1 = 10, \lambda_2 = 2, \lambda_3 = 0.15, \lambda_4 = 0.05$ and $\lambda_5 = 0.02$,

and that the remaining eigenvalues are all zero. Please work out which principal components we should use to project the dataset such that afterwards we can retain at least 95% of the variance of this dataset? Please show your numerical derivation in detail. **[4 marks]**

Do not complete the attendance slip, fill in the front of the answer book or turn over the question paper until you are told to do so

Important Reminders

- Coats/outwear should be placed in the designated area.
- Unauthorised materials (e.g. notes or Tippex) must be placed in the designated area.
- Check that you do not have any unauthorised materials with you (e.g. in your pockets, pencil case).
- Mobile phones and smart watches **must** be switched off and placed in the designated area or under your desk. They must not be left on your person or in your pockets.
- You are not permitted to use a mobile phone as a clock. If you have difficulty seeing a clock, please alert an Invigilator.
- You are not permitted to have writing on your hand, arm or other body part.
- Check that you do not have writing on your hand, arm or other body part – if you do, you must inform an Invigilator immediately
- Alert an Invigilator immediately if you find any unauthorised item upon you during the examination.

Any students found with non-permitted items upon their person during the examination, or who fail to comply with Examination rules may be subject to Student Conduct procedures.