

ISC Class 12 Artificial Intelligence Specimen Paper 1

Written by the Future Topper subject team on the CISCE examination pattern. Not an official CISCE paper.
Found a mistake? Write to contactfuturetopper@gmail.com

Maximum Marks: 70 | Time Allowed: Three hours | Reading Time: 15 minutes

Paper 1 (Theory) carries 70 marks. Part I is compulsory. In Part II attempt one question from Section A, two from Section B and two from Section C.

General Instructions

Write your answers on the separate answer paper provided. The first 15 minutes are for reading the paper, and you may not write during that time. Part I is compulsory and carries 20 marks. Part II carries 50 marks and is divided into three Sections. From Section A answer any one of the two questions, from Section B answer any two of the three questions, and from Section C answer any two of the three questions. Marks for each question, or part of a question, are shown in brackets. Where a question asks for code, write it in Python; you may use the NumPy, Pandas, scikit-learn, Keras (TensorFlow), Matplotlib and OpenCV libraries named in the syllabus. Show all working for every calculation. The numbers have been chosen so that the arithmetic can be done by hand, and a value may be given correct to three decimal places unless the question says otherwise. No graphs, network diagrams, images, data tables or confusion matrices are printed in this paper; all data are given in words. Where a question asks for a diagram you are expected to draw it yourself in pencil and label it. This paper carries 70 of the 100 marks for the subject, the remaining 30 being for Paper 2 (Practical), which is made up of the practical examination and the practical file.

PART I (20 Marks)

Answer all questions. Question 1 carries 10 marks and Question 2 carries 10 marks.

Question 1 [10]

Each part carries 1 mark. For the parts that offer four choices, write only the option you choose. For the rest, answer in a single sentence or give the single word, name or value asked for.

- (i) In the sentence "Sunita flew to Kolkata on Monday", a named entity recogniser would tag the word Kolkata as:
(a) a person (b) a location (c) a date (d) an organisation
- (ii) Each bulb made on a line is defective with probability 0.05, independently of the others. The number of defective bulbs in a box of 10 follows:
(a) a uniform distribution (b) a normal distribution
(c) a binomial distribution (d) an exponential distribution
- (iii) State what is meant by a Type I error in hypothesis testing.
- (iv) The activation function that passes a positive input through unchanged and turns every negative input into 0 is:
(a) ReLU (b) sigmoid (c) tanh (d) softmax
- (v) Name the layer of a convolutional neural network that shrinks the width and height of each feature map by keeping only the largest value in every small window.
- (vi) The scikit-learn function used to divide a data set into a training set and a test set is:
(a) fit_transform() (b) train_test_split()
(c) get_dummies() (d) predict()
- (vii) A network is trained on 2,000 samples with a batch size of 50. The number of times the weights are updated in one epoch is:
(a) 20 (b) 50 (c) 40 (d) 100

- (viii) Name the approach that flags a bank transaction as suspicious because it does not fit the normal pattern of a customer's spending, without needing examples labelled as fraud.
- (ix) Which of the following is a hyperparameter, chosen before training begins rather than learnt from the data?
 - (a) a weight of the first layer (b) the bias of the output neuron
 - (c) the gradient of the loss (d) the learning rate
- (x) When OpenCV reads a colour image with `cv2.imread()`, the three colour channels are stored in the order:
 - (a) BGR (b) RGB (c) HSV (d) CMYK

Question 2 [10]

Each part carries 2 marks.

- (i) A customer support chatbot received the following numbers of complaints on five days: 21, 23, 24, 25 and 27. Find the mean and the population standard deviation of the number of complaints. [2]
- (ii) A single artificial neuron receives the inputs 2, -1 and 3 with the weights 0.5, 0.4 and -0.2 respectively, and has a bias of 0.1. Find the weighted sum and the output of the neuron when its activation function is ReLU. State the output if the activation were instead a step function that gives 1 for a weighted sum greater than 0 and 0 otherwise. [2]
- (iii) A student scales the whole data set with a MinMaxScaler and only then splits it into a training set and a test set. State what is wrong with this order, and give the correct order of the steps. [2]
- (iv) A bank uses a model to check every card payment for fraud. Explain what a false positive and a false negative mean for this model, and state which of the two can make the customer lose money. [2]
- (v) Distinguish between image classification and object detection, giving one example of each. [2]

PART II

This Part carries 50 marks. Answer one question from Section A, two questions from Section B and two questions from Section C.

SECTION A (10 Marks)

Answer any one question from this Section. Each question carries 10 marks. This Section covers applications of AI: language processing, statistics and probability.

Question 3 [10]

- (a) Read the sentence: "On 15 August 2026, Priya Raman of Infosys opened a new office in Mysuru worth 300 crore rupees." List every named entity in the sentence with the entity type a named entity recogniser would give it. Then state one practical use of named entity recognition. [3]
- (b) A spam filter is built from past messages. Of all messages, 30 per cent are spam. The word "lottery" appears in 40 per cent of spam messages and in 2 per cent of genuine messages. Using Bayes' theorem, find the probability that a new message containing "lottery" is spam. Show your working, and explain in one sentence how a Naive Bayes filter extends this idea to a whole message. [4]
- (c) Explain how extractive text summarisation produces a summary, setting out its main steps, and state one difference between extractive and abstractive summarisation. [3]

Question 4 [10]

- (a) Five students recorded the hours x they spent each week on an AI practice platform and their marks y in a test: (2, 38), (4, 48), (5, 58), (6, 78) and (8, 78). Find the least squares regression line of y on x , and use it to estimate the marks of a student who practises for 7 hours a week. [4]
- (b) For the same data, find Karl Pearson's correlation coefficient r and interpret its value. [3]

- (c) An ed-tech company claims that students who use its AI tutor score, on average, more than 60 in a standard test. A researcher sets out to test the claim. State suitable null and alternative hypotheses, and describe a Type I error and a Type II error in the context of this test. [3]

SECTION B (20 Marks)

Answer any two questions from this Section. Each question carries 10 marks. This Section covers neural networks, machine learning and deep learning, and building and evaluating them in Python.

Question 5 [10]

- (a) A small neural network has two inputs x_1 and x_2 , one hidden layer of three neurons that use the ReLU activation, and one output neuron with no activation (a linear output). Hidden neuron 1 has weights 0.5 from x_1 and 0.25 from x_2 and bias 0. Hidden neuron 2 has weights -0.5 and 0.75 and bias -0.5. Hidden neuron 3 has weights 0.4 and -0.6 and bias 0.2. The output neuron has weights $v_1 = 0.8$, $v_2 = 0.6$ and $v_3 = 0.5$ from hidden neurons 1, 2 and 3, and bias 0.1. Draw the network yourself and label every layer. Then carry out a forward pass for the input $x_1 = 1$, $x_2 = 2$, finding the output of each hidden neuron and the predicted output. [4]
- (b) The target output for this input is 2. Using the loss $L = (1/2)(y - \hat{y})^2$, find the loss. The gradient of L with respect to an output weight v_j is $(\hat{y} - y)h_j$, the error multiplied by h_j , and with respect to the output bias it is $(\hat{y} - y)$. Carry out one step of gradient descent with learning rate 0.1 on v_1 , v_2 , v_3 and the output bias. Explain why v_3 does not change, and show that the new prediction is closer to the target. [4]
- (c) A multi-layer perceptron is to recognise handwritten digits 0 to 9 from grey scale images of 28 by 28 pixels. State the number of neurons needed in the input layer and in the output layer, and name the activation function used in the output layer, with a reason. [2]

Question 6 [10]

- (a) Give the output of the following program. Then find the min-max scaled value of the area of the flat in the third row (index 2), correct to three decimal places. [4]

```
import pandas as pd
df = pd.DataFrame({
    "city": ["Pune", "Delhi", "Pune", "Chennai"],
    "area": [650, 1200, 900, 1500],
    "price": [52, 96, 70, 110]})
enc = pd.get_dummies(df, columns=["city"], dtype=int)
print(enc.shape)
print(list(enc.columns))
print(enc.loc[1].tolist())
```

- (b) Study the following Keras code for a model that predicts whether a loan applicant will repay, from 8 input features.

```
from tensorflow import keras
model = keras.Sequential([
    keras.Input(shape=(8,)),
    keras.layers.Dense(16, activation="relu"),
    keras.layers.Dense(8, activation="relu"),
    keras.layers.Dense(1, activation="sigmoid")
])
model.compile(optimizer="adam", loss="binary_crossentropy",
    metrics=["accuracy"])
```

State the number of trainable parameters in each Dense layer and in the whole model, showing how you obtain them. Give a reason for using a sigmoid output with binary cross-entropy loss here. [3]

- (c) The data set has 1,600 rows and is first split 80:20 into training and test sets. The model is then trained with `model.fit(X_train, y_train, epochs=20, batch_size=32, validation_split=0.2)`. State how many rows are actually used to update the weights, how many batches make one epoch, and how many weight updates take place in all. Explain what an epoch is. [3]

Question 7 [10]

- (a) A model checks photographs of tomato leaves taken by farmers in Nashik and outputs 1 for a diseased leaf and 0 for a healthy one. Give the output of the following program, which compares its predictions with the true labels for twelve leaves, and state in words what the precision and recall figures mean here. [4]

```
y_true = [1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0]
y_pred = [1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0]
tp = fp = fn = tn = 0
for t, p in zip(y_true, y_pred):
    if t == 1 and p == 1:
        tp += 1
    elif t == 0 and p == 1:
        fp += 1
    elif t == 1 and p == 0:
        fn += 1
    else:
        tn += 1
print(tp, fp, fn, tn)
print(round((tp + tn) / len(y_true), 3))
print(round(tp / (tp + fp), 3), round(tp / (tp + fn), 3))
```

- (b) A grid search is used to tune a neural network. The learning rate is tried at 0.1, 0.01 and 0.001, the batch size at 16, 32 and 64, and the number of hidden layers at 1 and 2, and every setting is scored by 5-fold cross-validation. Find how many combinations are tried and how many times a model is trained. If a random search with 10 sampled combinations is used instead, how many trainings are needed? Give one advantage of each method. [3]
- (c) After training, the object history returned by `fit()` holds the lists `history.history["accuracy"]` and `history.history["val_accuracy"]`. Write Python statements using Matplotlib to plot both against the epoch number on one graph, with axis labels, a title and a legend. State what it means if the gap between the two curves keeps widening. [3]

SECTION C (20 Marks)

Answer any two questions from this Section. Each question carries 10 marks. This Section covers practical uses of machine learning and computer vision.

Question 8 [10]

- (a) A textile mill in Surat runs 40 power looms. Each loom has sensors that log vibration and motor temperature every hour, and the mill keeps a record of every breakdown. Explain how the mill can frame predictive maintenance as a supervised classification problem, stating what one row of data, the features and the label would be. Name the regression problem that could be set instead. [3]
- (b) A sample of 20 loom-weeks, chosen from the mill's records to include plenty of failures, shows 8 failures in the following week and 12 with no failure. Splitting on vibration gives 10 loom-weeks with high vibration (7 failed, 3 did not) and 10 with low vibration (1 failed, 9 did not). Splitting on temperature gives 8 with high temperature (4 failed, 4 did not) and 12 with normal temperature (4 failed, 8 did not). The Gini impurity of a group is 1 minus the sum of the squares of the proportions of each class in it, and the weighted impurity of a split is the average of the impurities of its branches, each weighted by the fraction of rows it holds. A decision tree splits on the feature that gives the lowest weighted impurity. Find the impurity before any split and the weighted impurity after each split, and decide which feature the tree should use at its root. [4]
- (c) Breakdowns are rare: in the mill's full records only 3 loom-weeks in every 100 are followed by a failure. Explain why a model with 97 per cent accuracy may be useless here, state which kind of error is costlier for the mill, and name the metric it should watch most closely. [3]

Question 9 [10]

- (a) Machine learning is used in medical diagnosis with three broad kinds of data: medical images, genetic information and patient records. For each kind, give one way in which a model can use it. [3]
- (b) A district hospital in Odisha uses a model to screen chest X-rays for tuberculosis. The model has a recall (sensitivity) of 95 per cent and a specificity of 90 per cent, and 2 per cent of the people screened actually have tuberculosis. For 10,000 people screened, find the numbers of true positives, false negatives, false positives and true negatives. Hence find the fraction of people flagged positive who really have tuberculosis, and comment on what the hospital should do with a positive result. [4]
- (c) Give two reasons why the final decision on a diagnosis should stay with a doctor even when the model is accurate, and state one step the hospital should take to protect patients' data. [3]

Question 10 [10]

- (a) Give the output of the following program. Explain why the green square, and not the red one, passes the `cv2.inRange()` test. [4]

```
import numpy as np
import cv2
img = np.zeros((60, 80, 3), np.uint8)
img[10:30, 10:40] = (0, 255, 0)
img[35:55, 50:70] = (0, 0, 255)
hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)
mask = cv2.inRange(hsv, (40, 100, 100), (80, 255, 255))
print(img.shape)
print(hsv[20, 20], hsv[40, 60])
print(cv2.countNonZero(mask))
```

- (b) An object detection model at a toll plaza predicts a bounding box for a car with top-left corner (40, 30) and bottom-right corner (120, 90), in pixels. The true box, marked by a person, has corners (80, 50) and (160, 110). Find the intersection over union (IoU) of the two boxes. If a detection counts as correct only when IoU is at least 0.5, is this detection correct? [3]
- (c) A college plans a facial recognition attendance system using a classroom camera. Describe the main stages from a camera frame to a marked attendance register, and state one safeguard the college should adopt for students' privacy. [3]

ANSWER KEY AND MARKING SCHEME

A full model answer is given for every part. Marks shown in italics are the step marks an examiner would award. Every calculation in this key has been checked by machine, and every Python program quoted in a question has been run with Python 3, NumPy, Pandas, scikit-learn and OpenCV, so the outputs given are the outputs actually produced. The one exception is Keras code, which is marked on its structure, with the parameter counts checked by arithmetic. Accept any correct method that reaches the same value, and accept a value rounded sensibly. For an output question the values and their order must be right, but differences in the spacing NumPy prints inside square brackets are not penalised. For a question that asks for code, accept any sensible variable names and any equivalent library call. For an explanation, credit the idea rather than the exact wording, and accept any relevant Indian or real world example. For a drawing accept any clear and correctly labelled version.

Question 1 [10]

- (i) (b) a location. Kolkata is the name of a place, so it is tagged as a location (GPE, a geo-political entity, in spaCy). Sunita is a person and Monday is a date.
- (ii) (c) a binomial distribution. There is a fixed number of independent trials (10 bulbs), each with two outcomes and the same probability of "success", here 0.05, and we count the successes.
- (iii) A Type I error is rejecting the null hypothesis when it is in fact true, a false alarm. Its probability is the significance level, alpha.
- (iv) (a) ReLU, the rectified linear unit, $f(z) = \max(0, z)$.
- (v) The max pooling layer (a pooling layer). A 2×2 max pool with stride 2 halves the width and the height.
- (vi) (b) `train_test_split()`, from `sklearn.model_selection`.
- (vii) (c) 40. One update is made per batch, and $2,000 / 50 = 40$ batches make one epoch.
- (viii) Anomaly detection (outlier detection), an unsupervised approach.
- (ix) (d) the learning rate. Weights and biases are parameters learnt during training, and a gradient is computed during training; the learning rate is set by the programmer beforehand.
- (x) (a) BGR, that is blue, green, red. This is why an image must be converted with `cv2.COLOR_BGR2RGB` before it is shown with Matplotlib.

1 mark for each correct part. No mark is deducted for a wrong option, and a one-line answer that carries the correct idea gains the mark even if it is loosely worded.

Question 2 [10]

- (i) Mean = $(21 + 23 + 24 + 25 + 27) / 5 = 120 / 5 = 24$. Deviations from the mean are -3, -1, 0, 1, 3, whose squares are 9, 1, 0, 1, 9, adding to 20. Population variance = $20 / 5 = 4$, so the standard deviation = square root of 4 = 2 complaints.

Note for the examiner: A candidate who divides by $n - 1$ and gives the sample standard deviation, square root of 5 = 2.236, earns the method mark only and not the answer mark.

Mean 24 with the squared deviations totalling 20 (method) 1. Population variance 4 and standard deviation 2 (answer) 1.

- (ii) Weighted sum $z = (2)(0.5) + (-1)(0.4) + (3)(-0.2) + 0.1 = 1.0 - 0.4 - 0.6 + 0.1 = 0.1$. ReLU output = $\max(0, 0.1) = 0.1$. With the step function, $z = 0.1$ is greater than 0, so the output is 1.

Weighted sum 0.1 including the bias 1. ReLU output 0.1 and step output 1 1.

- (iii) The scaler has learnt the minimum and maximum from the test rows as well, so information from the test set has leaked into training. The test score is then optimistic and no longer an honest measure of performance on unseen data. Correct order: split first; fit the scaler on the training set only (`fit_transform` on `X_train`); then apply the same fitted scaler to the test set (`transform` on `X_test`).

Data leakage from the test set explained 1. Split, fit on training only, transform test 1.

- (iv) A false positive is a genuine payment that the model wrongly flags as fraud, so it is declined or held; the customer is inconvenienced, for example at a shop counter. A false negative is a fraudulent payment that the model wrongly passes as genuine, so the money leaves the account. The false negative is the one that makes the customer lose money.

False positive and false negative both described in context 1. False negative identified as the one that loses money 1.

- (v) Image classification gives one label to the whole image, for example deciding whether a photo of a leaf shows a healthy or a diseased leaf. Object detection finds each object in the image, gives its class and draws a bounding box round it, so it answers both what and where; for example locating every car and two-wheeler in a traffic camera frame.

Classification: one label for whole image, with example 1. Detection: class plus bounding box location for each object, with example 1.

Question 3 [10]

- (a) Entities: "15 August 2026" is a DATE; "Priya Raman" is a PERSON; "Infosys" is an ORGANISATION (ORG); "Mysuru" is a location (GPE); "300 crore rupees" is MONEY. One use: a news or legal document system extracts the people, companies and places mentioned in thousands of articles so that they can be indexed, searched and linked; a bank can also pull names, dates and amounts out of loan documents automatically.

Note for the examiner: Accept LOC for Mysuru, and accept CARDINAL or QUANTITY for "300 crore" if "rupees" is not included, since recognisers differ.

Four or five entities with correct types (four for 1 mark, all five for 2) 2. One practical use 1.

- (b) Let S be spam and L the event that the message contains "lottery". $P(S) = 0.3$, $P(\text{not } S) = 0.7$, $P(L | S) = 0.4$, $P(L | \text{not } S) = 0.02$. By Bayes' theorem $P(S | L) = P(L | S) P(S) / [P(L | S) P(S) + P(L | \text{not } S) P(\text{not } S)] = (0.4 \times 0.3) / (0.4 \times 0.3 + 0.02 \times 0.7) = 0.12 / (0.12 + 0.014) = 0.12 / 0.134 = 60/67 = 0.896$ (about 89.6 per cent). Equivalently, out of 1,000 messages 300 are spam and 120 of them contain the word, 700 are genuine and 14 of them contain it, so 120 of the 134 messages with the word are spam. A Naive Bayes filter multiplies the likelihoods of all the words in the message, assuming the words occur independently given the class, and labels the message spam if the resulting spam probability is the larger.

Note for the examiner: Value checked by machine as $60/67 = 0.8955$. Accept 0.895 or 0.9.

Bayes' theorem written with the correct terms 1. Substitution 0.12 and 0.014 1. Answer 0.896 1. Naive Bayes: product of word likelihoods under the independence assumption 1.

- (c) Steps: the text is split into sentences and words; stop words such as "the" and "is" are removed and the words may be lower-cased or stemmed; each remaining word is given a weight, usually its frequency (or TF-IDF); each sentence is scored by adding the weights of its words; the highest scoring sentences are picked and printed in their original order as the summary. Difference: an extractive summary copies whole sentences from the original text, while an abstractive summary generates new sentences in its own words, as a large language model does, and so may contain words that are not in the original.

Splitting and stop word removal 1. Word weights, sentence scores and top sentences chosen 1. Difference from abstractive 1.

Question 4 [10]

- (a) Mean of $x = 25 / 5 = 5$ and mean of $y = 300 / 5 = 60$. Deviations of x : -3, -1, 0, 1, 3; deviations of y : -22, -12, -2, 18, 18. $S_{xx} = 9 + 1 + 0 + 1 + 9 = 20$. $S_{xy} = (-3)(-22) + (-1)(-12) + (0)(-2) + (1)(18) + (3)(18) = 66 + 12 + 0 + 18 + 54 = 150$. Slope $b = S_{xy} / S_{xx} = 150 / 20 = 7.5$. Intercept $a = 60 - 7.5 \times 5 = 22.5$. Regression line: $y = 22.5 + 7.5x$. For $x = 7$: $y = 22.5 + 52.5 = 75$ marks.

Note for the examiner: Machine checked. A candidate using the formula $b = (n \sum xy - \sum x \sum y) / (n \sum x^2 - (\sum x)^2)$ gets $(5 \times 1650 - 25 \times 300) / (5 \times 145 - 625) = 750 / 100 = 7.5$, which is equally acceptable.

Means 5 and 60 1. $S_{xx} = 20$ and $S_{xy} = 150$ 1. Line $y = 22.5 + 7.5x$ 1. Estimate 75 1.

- (b) $S_{yy} = 484 + 144 + 4 + 324 + 324 = 1280$. $r = S_{xy} / \text{square root of } (S_{xx} \times S_{yy}) = 150 / \text{square root of } (20 \times 1280) = 150 / \text{square root of } 25600 = 150 / 160 = 0.9375$. Interpretation: r is close to +1, so there is a strong positive linear relationship; students who practise longer tend to score higher. It does not by itself prove that practice causes the higher marks.

$S_{yy} = 1280$ 1. $r = 0.9375$ (0.94) 1. Strong positive linear relationship interpreted 1.

- (c) H_0 : the mean score of AI tutor users is 60 ($\mu = 60$, or μ at most 60). H_1 : the mean score is greater than 60 ($\mu > 60$). Type I error: the researcher concludes that the tutor raises the mean above 60 when it really does not, so schools may pay for a product that does not work. Type II error: the researcher fails to reject H_0 and concludes there is no evidence of improvement when the tutor really does raise the mean above 60, so a useful product is wrongly dismissed.

H_0 and H_1 correctly stated 1. Type I error in context 1. Type II error in context 1.

Question 5 [10]

- (a) Diagram: two input nodes, three hidden nodes, one output node, each input joined to every hidden node and every hidden node joined to the output, with the layers labelled input, hidden and output. Forward pass. Hidden 1: $z_1 = 0.5(1) + 0.25(2) + 0 = 1.0$, $h_1 = \text{ReLU}(1.0) = 1.0$. Hidden 2: $z_2 = -0.5(1) + 0.75(2) - 0.5 = 0.5$, $h_2 = 0.5$. Hidden 3: $z_3 = 0.4(1) - 0.6(2) + 0.2 = -0.6$, $h_3 = \text{ReLU}(-0.6) = 0$. Output: $\hat{y} = 0.8(1.0) + 0.6(0.5) + 0.5(0) + 0.1 = 0.8 + 0.3 + 0 + 0.1 = 1.2$.

Network drawn and labelled with layers and connections 1. $h_1 = 1.0$ and $h_2 = 0.5$ 1. $h_3 = 0$ with ReLU applied to -0.6 1. Predicted output 1.2 1.

- (b) Loss $L = 0.5 \times (2 - 1.2)^2 = 0.5 \times 0.64 = 0.32$. Error term $\hat{y} - y = 1.2 - 2 = -0.8$. Gradients: for v_1 , $-0.8 \times 1.0 = -0.8$; for v_2 , $-0.8 \times 0.5 = -0.4$; for v_3 , $-0.8 \times 0 = 0$; for the bias, -0.8 . Update rule: new value = old value - 0.1 $\times$ gradient. $v_1 = 0.8 + 0.08 = 0.88$; $v_2 = 0.6 + 0.04 = 0.64$; $v_3 = 0.5 - 0 = 0.5$; bias = $0.1 + 0.08 = 0.18$. v_3 does not change because hidden neuron 3 gave an output of 0 (ReLU cut off its negative sum), so it made no contribution to this prediction and its gradient is zero. New prediction = $0.88(1.0) + 0.64(0.5) + 0.5(0) + 0.18 = 1.38$, and the new loss is $0.5 \times (0.62)^2 = 0.1922$, which is less than 0.32. This is one step of backward propagation for the output layer.

Note for the examiner: All values machine checked.

Loss 0.32 1. Gradients -0.8, -0.4, 0 and -0.8 1. Updated values 0.88, 0.64, 0.5 and 0.18 1. Reason v_3 unchanged and new prediction 1.38 1.

- (c) Input layer: $28 \times 28 = 784$ neurons, one for each pixel after the image is flattened. Output layer: 10 neurons, one for each digit. Output activation: softmax, because it turns the 10 outputs into probabilities that add up to 1, and the digit with the largest probability is the prediction.

784 input and 10 output neurons 1. Softmax with reason 1.

Question 6 [10]

- (a) Output, as produced by running the program:

```
(4, 5)
['area', 'price', 'city_Chennai', 'city_Delhi', 'city_Pune']
[1200, 96, 0, 1, 0]
```

Reasoning: get_dummies replaces the city column by one 0/1 column for each distinct city, named city_ followed by the city and placed after the other columns in alphabetical order (Chennai, Delhi, Pune), so there are 4 rows and $2 + 3 = 5$ columns. Row 1 is the Delhi flat, so its city columns are 0, 1, 0. Min-max scaling: $x' = (x - \min) / (\max - \min) = (900 - 650) / (1500 - 650) = 250 / 850 = 0.294$.

Note for the examiner: Accept the column names without quotes or brackets provided the order is right.

Shape (4, 5) 1. Column list in the right order 1. Row values [1200, 96, 0, 1, 0] 1. Scaled value 0.294 with formula 1.

- (b) Trainable parameters of a Dense layer = (inputs $\times$ neurons) + neurons, one weight per connection and one bias per neuron. First hidden layer: $8 \times 16 + 16 = 144$. Second hidden layer: $16 \times 8 + 8 = 136$. Output layer: $8 \times 1 + 1 = 9$. Total = $144 + 136 + 9 = 289$. The output is a yes or no answer, so one sigmoid neuron gives a probability between 0 and 1 of repayment, and binary cross-entropy is the loss that measures the error of such a probability against a 0 or 1 label.

Note for the examiner: Parameter counts checked by arithmetic.

Formula and 144, 136, 9 1. Total 289 1. Reason for sigmoid with binary cross-entropy 1.

- (c) Training set = 80 per cent of 1,600 = 1,280 rows. validation_split = 0.2 holds back the last 20 per cent of these, 256 rows, for validation, so 1,024 rows update the weights. Batches per epoch = $1,024 / 32 = 32$. Total

updates = $32 \times 20 = 640$. An epoch is one complete pass of the whole training data through the network, forward and backward, so that every training row has been used once.

1,024 rows after the validation split 1. 32 batches and 640 updates 1. Epoch explained 1.

Question 7 [10]

(a) Output, as produced by running the program:

```
4 1 2 5
0.75
0.8 0.667
```

Working: comparing the lists position by position gives true positives at positions 1, 4, 6 and 9 ($tp = 4$), a false positive at position 7 ($fp = 1$), false negatives at positions 3 and 11 ($fn = 2$) and true negatives at the other five positions ($tn = 5$). Accuracy = $9 / 12 = 0.75$. Precision = $4 / 5 = 0.8$. Recall = $4 / 6 = 0.667$. Meaning: precision 0.8 says that 80 per cent of the leaves the model called diseased really were diseased; recall 0.667 says that the model found only two thirds of the diseased leaves and missed one in three.

4 1 2 5 1. 0.75 1. 0.8 0.667 1. Meaning of precision and recall in context 1.

(b) Combinations = $3 \times 3 \times 2 = 18$. With 5-fold cross-validation each is trained 5 times, so $18 \times 5 = 90$ trainings. Random search with 10 combinations needs $10 \times 5 = 50$ trainings. Advantage of grid search: it is exhaustive, so the best combination on the grid is certain to be tried. Advantage of random search: it costs less time and, for the same budget, explores more different values of each hyperparameter, which matters when only one or two hyperparameters really affect the result.

18 combinations and 90 trainings 1. 50 trainings for random search 1. One advantage of each 1.

(c) One correct set of statements:

```
import matplotlib.pyplot as plt
acc = history.history["accuracy"]
val = history.history["val_accuracy"]
epochs = range(1, len(acc) + 1)
plt.plot(epochs, acc, label="Training accuracy")
plt.plot(epochs, val, label="Validation accuracy")
plt.xlabel("Epoch")
plt.ylabel("Accuracy")
plt.title("Training and validation accuracy")
plt.legend()
plt.show()
```

If training accuracy keeps rising while validation accuracy levels off or falls, so that the gap widens, the model is overfitting: it is memorising the training data rather than learning patterns that carry over to new data.

Note for the examiner: The plotting code has been run against a stand-in history object. Accept `plt.plot(acc)` without an explicit epoch range.

Both lists plotted against epoch with labels for the legend 1. Axis labels, title, legend and show 1. Widening gap identified as overfitting 1.

Question 8 [10]

(a) One row is one loom over one period, say a loom-week. Features: summary values of the sensor logs for that week, such as average and peak vibration, average temperature and its rise, running hours, days since the last service and age of the loom. Label: 1 if the loom broke down in the following week and 0 if it did not, taken from the breakdown records. A classifier trained on past loom-weeks then flags looms likely to fail next week. Regression alternative: predict the remaining useful life, the number of days or hours left before the next failure.

Row and features 1. Label defined from future breakdowns 1. Remaining useful life as regression 1.

(b) Gini = $1 - \text{sum of } p^2$. Root: $1 - (0.4^2 + 0.6^2) = 1 - (0.16 + 0.36) = 0.48$. Vibration: high, $1 - (0.7^2 + 0.3^2) = 1 - 0.58 = 0.42$; low, $1 - (0.1^2 + 0.9^2) = 1 - 0.82 = 0.18$; weighted = $(10/20)(0.42) + (10/20)(0.18) = 0.30$; reduction = $0.48 - 0.30 = 0.18$. Temperature: high, $1 - (0.5^2 + 0.5^2) = 0.5$; normal, $1 - ((1/3)^2 + (2/3)^2) = 1 - 5/9$

$= 4/9 = 0.444$; weighted $= (8/20)(0.5) + (12/20)(0.444) = 0.2 + 0.267 = 0.467$; reduction $= 0.48 - 0.467 = 0.013$. The vibration split leaves the lower weighted impurity (the larger Gini gain), so the tree should split on vibration at the root.

Note for the examiner: Values machine checked with exact fractions ($0.48 = 12/25$, $0.30 = 3/10$, $0.467 = 7/15$).

Root Gini 0.48 1. Vibration branches and weighted 0.30 1. Temperature branches and weighted 0.467 1. Vibration chosen with reason 1.

- (c) A model that always predicts "no failure" is right 97 times in 100, so it scores 97 per cent accuracy while catching no failures at all; accuracy hides the minority class. A false negative, a loom that breaks down without warning, is costlier, because it stops production, may damage cloth and parts and needs emergency repair, whereas a false positive costs only an early inspection. The mill should watch recall on the failure class (with precision kept reasonable, for example through the F1 score).

Always-no model gives 97 per cent, so accuracy misleads 1. False negative costlier, with reason 1. Recall named 1.

Question 9 [10]

- (a) Medical images: a convolutional network studies X-rays, CT or MRI scans or retina photographs and flags signs of disease such as tuberculosis, tumours or diabetic eye damage for a doctor to check. Genetic information: a model looks for gene variants linked with a disease, so as to estimate a person's risk of, say, an inherited cancer or to suggest which medicine is likely to work for that patient. Patient records: a model uses age, blood test results, blood pressure, history and symptoms to predict risks such as diabetes or heart disease, or to warn when a patient in hospital is getting worse.

Medical images use 1. Genetic information use 1. Patient records use 1.

- (b) People with TB = 2 per cent of 10,000 = 200; without TB = 9,800. True positives = 95 per cent of 200 = 190; false negatives = $200 - 190 = 10$. True negatives = 90 per cent of 9,800 = 8,820; false positives = $9,800 - 8,820 = 980$. Flagged positive = $190 + 980 = 1,170$, of whom 190 have TB: $190 / 1,170 = 0.162$, about 16 per cent (this is the precision, or positive predictive value). Comment: because the disease is rare, most positive flags are false alarms, so a positive result must be confirmed by a doctor and a sputum or molecular test before treatment; the model is useful as a screen because it misses only 10 of the 200 cases.

Note for the examiner: Machine checked: $190/1170 = 19/117 = 0.1624$. The overall accuracy, $9,010/10,000 = 90.1$ per cent, may be quoted but is not required.

TP 190 and FN 10 1. TN 8,820 and FP 980 1. Fraction 0.162 1. Comment: confirmatory test, screen not diagnosis 1.

- (c) Reasons: the model can be wrong in ways that are hard to see, for example on patients unlike those in its training data, such as a different age group, X-ray machine or region, so its errors are not spread fairly; a deep network does not explain its reasoning, while a doctor can combine the scan with symptoms and history and is accountable for the decision. Data step: remove names and identifiers (anonymise) before images are used for training, store them encrypted with access limited to authorised staff, and take patients' informed consent.

First reason 1. Second reason 1. One data protection step 1.

Question 10 [10]

- (a) Output, as produced by running the program:

```
(60, 80, 3)
[ 60 255 255] [ 0 255 255]
600
```

Reasoning: the image is 60 rows by 80 columns with 3 channels. OpenCV stores BGR, so (0, 255, 0) is pure green and (0, 0, 255) is pure red. In OpenCV the hue runs from 0 to 179 (half of the usual 0 to 360 degrees), so pure green, 120 degrees, has hue 60, and pure red has hue 0; both have full saturation and value, 255. The pixel at row 20, column 20 lies in the green square and the pixel at row 40, column 60 lies in the red square. The mask keeps pixels whose hue is between 40 and 80, so only the green square passes: 20 rows x 30 columns = 600 white pixels. The red square, with hue 0, is outside the range.

(60, 80, 3) 1. HSV values [60 255 255] and [0 255 255] 1. 600 1. Reason: green hue 60 lies in 40 to 80, red hue 0 does not 1.

- (b)** Overlap: x from $\max(40, 80) = 80$ to $\min(120, 160) = 120$, width 40; y from $\max(30, 50) = 50$ to $\min(90, 110) = 90$, height 40. Intersection = $40 \times 40 = 1,600$. Area of each box = $80 \times 60 = 4,800$. Union = $4,800 + 4,800 - 1,600 = 8,000$. IoU = $1,600 / 8,000 = 0.2$. Since 0.2 is less than 0.5, the detection is not counted as correct; the box is badly placed even though it touches the car.

Note for the examiner: Machine checked.

Intersection 1,600 1. Union 8,000 and IoU 0.2 1. Not correct, since $0.2 < 0.5$ 1.

- (c)** Stages: (1) capture frames from the camera with cv2.VideoCapture; (2) face detection, finding each face and its bounding box, for example with a Haar cascade or a deep learning detector; (3) face recognition, where each detected face is turned into an embedding, a vector of numbers, and compared with the stored embeddings of enrolled students; the closest match within a threshold gives the student's name; (4) the name and time are written to the attendance file, once per student per lecture. Safeguard: take written consent, store only the embeddings rather than photographs, encrypt them, use them only for attendance and delete them when the student leaves; offer a manual alternative for those who opt out.

Capture and detection 1. Recognition by matching embeddings and marking the register 1. One privacy safeguard 1.