

THEME A: CONCEPTS · A4.3 · HL ONLY

Machine learning approaches

Original practice questions · 34 marks · about 55 minutes · covers A4.3.1–A4.3.10

Instructions. Answer all questions. The maximum mark for this paper is [34 marks]. The marks available are shown in brackets. Parts badged **HL** assess higher level only content; standard level students may leave these out. **A4.3 is a higher level only topic**, so every question in this paper is HL.

1

Total: 6 marks

A city energy provider fits a linear regression model to predict a household's monthly electricity consumption from the floor area of the property. The fitted model is shown below.

$$\text{predicted monthly consumption (kWh)} = 1.8 \times \text{floor area (m}^2\text{)} + 55$$
$$\text{coefficient of determination, } r^2 = 0.88$$

- (a) **HL** Identify the independent variable and the dependent variable in this model. [2]

- (b) **HL** State what the value 1.8 tells the provider. [2]

- (c) **HL** Outline what the r^2 value of 0.88 indicates about the model. [2]

The provider also wants to predict whether a household will fall into arrears. Each household is labelled either arrears or no arrears.

- (a) **HL** Describe how the K-Nearest Neighbours (K-NN) algorithm would classify a new household. [3]

- (b) **HL** Outline **one** real-world application of a decision tree. [2]

- (c) **HL** State **one** reason the provider might choose a decision tree rather than a neural network. [1]

Only 2 per cent of the provider's households are in arrears. A first model predicts no arrears for every household and reports 98 per cent accuracy.

- (a) **HL** Distinguish between overfitting and underfitting. [3]

- (b) **HL** Explain the role of hyperparameter tuning in addressing these problems. [2]

- (c) **HL** Outline why accuracy is a poor measure of this model's usefulness, and name a metric that would be more informative. [2]

4

Total: 5 marks

The provider holds unlabelled data on when and how much electricity each household uses.

- (a) **HL** Describe how clustering could help the provider group its customers. [3]

- (b) **HL** Outline what association rule discovery could reveal about how households use appliances. [2]

5

Total: 5 marks

The provider tests a self-improving system that schedules its repair engineers, and separately uses a genetic algorithm to plan the shortest route between the jobs on an engineer's list.

- (a) **HL** Outline the exploration versus exploitation trade-off in reinforcement learning. [3]

- (b) **HL** Identify **two** operators a genetic algorithm uses to produce each new generation. [2]

(a) **HL** Sketch a single perceptron. Label its inputs, weights, bias, activation function and output. **[3]**

(b) **HL** Outline **one** reason a convolutional neural network (CNN) is better suited than a multi- [2]

END OF QUESTIONS