

THEME A: CONCEPTS · A4.3 · HL ONLY

Machine learning approaches — Mark scheme

34 marks · covers A4.3.1–A4.3.10

How to mark. Each creditworthy point is on its own line and ends with a semicolon (;). A line beginning with ... is an *expansion* of the line above it, so the pair is worth the marks shown. A slash (/) separates acceptable alternative wordings. Words in (brackets) are not needed for the mark. Mark positively: credit what the student has got right rather than penalising what is missing, and accept any valid alternative not listed here. Do not award more than the maximum for a part. Where a later part uses a wrong answer from an earlier part correctly, award follow-through marks (FT). **HL** marks parts that assess higher level only content.

Q	ANSWER	MARKS
HL 1(a) A4.3.1	Award [2] max Award [1] for each. Independent (predictor) variable: floor area; Dependent (response) variable: monthly electricity consumption;	2
HL 1(b) A4.3.1	Award [2] max Award [1] for identifying it as the slope/gradient and [1] for interpreting it in context. 1.8 is the slope (gradient) of the regression line; ... (so) each additional square metre of floor area is associated with an extra 1.8 kWh of predicted monthly consumption; Accept a correct interpretation using a larger increment, for example 180 kWh for every extra 100 m ² . Do not credit an answer that describes it as the starting value, which is the intercept of 55.	2
HL 1(c) A4.3.1	Award [2] max Award [1] per point up to [2] max. r^2 measures how much of the variation in consumption the model accounts for; A value of 0.88 means 88 per cent of the variation is explained by floor area, so the model fits the data well; The remaining 12 per cent is due to other factors the model does not include, such as household size or heating type; Do not credit 'the model is 88 per cent accurate', which confuses r^2 with a classification metric.	2
HL 2(a) A4.3.2	Award [3] max Award [1] per point up to [3] max. The new household is plotted in the same feature space as the labelled households; The distance from it to every labelled household is calculated; The K closest labelled households (its nearest neighbours) are selected; The new household is given whichever label is most common among those K neighbours (a majority vote); The value chosen for K affects the result: too small and it is sensitive to noise, too large and local detail is lost;	3
HL 2(b) A4.3.2	Award [2] max Award [1] for naming a valid application and [1] for outlining it. Any one. Medical diagnosis;	2

Q	ANSWER	MARKS
	...(where) a patient's symptoms are followed down a series of branching tests to reach a likely condition; Credit or loan approval; ...(where) attributes such as income and repayment history determine the branch taken to an approve or refuse decision;	
HL 2(c) A4.3.10	Award [1] max Any one, [1] max. A decision tree is interpretable: the provider can see and explain the rules that produced a decision, which a neural network cannot easily provide; It trains far faster and needs much less data than a neural network; It needs less computing power to train and to run; It handles a mixture of numeric and categorical attributes without extensive preprocessing;	1
HL 3(a) A4.3.3	Award [3] max Award [1] per point up to [3] max. Both must be addressed. Overfitting is when a model learns the training data too closely, including its noise and its individual quirks; ...(so) it performs very well on the training data but poorly on data it has not seen; Underfitting is when a model is too simple to capture the underlying pattern in the data; ...(so) it performs poorly on the training data and on new data alike; The aim is a model that generalises: complex enough to capture the real pattern, but not so complex that it memorises the sample;	3
HL 3(b) A4.3.3	Award [2] max Award [1] per point up to [2] max. Hyperparameters are settings chosen before training, such as the value of K, tree depth or learning rate, rather than learned from the data; Tuning means trying different values and comparing performance on validation data to find the best combination; ...(so) the model's complexity can be increased to cure underfitting, or constrained to cure overfitting;	2
HL 3(c) A4.3.3	Award [2] max Award [1] for explaining why accuracy misleads here and [1] for naming a suitable metric. Because only 2 per cent of households are in arrears, a model that always predicts no arrears is right 98 per cent of the time while identifying none of the households that actually need help; ...(so) high accuracy hides the fact that the model has no useful predictive power on the class the provider cares about; A better metric: recall (the proportion of households genuinely in arrears that the model finds), precision, or the F1 score that combines the two; Award the second mark for any of recall, precision or F1. Naming a metric without explaining why accuracy misleads gains [1] max.	2
HL 4(a) A4.3.4	Award [3] max Award [1] per point up to [3] max. Must be applied to the provider. Clustering is unsupervised, so it works on the unlabelled usage data with no categories defined in advance;	3

Q	ANSWER	MARKS
	It groups households whose consumption patterns are similar to each other and different from those in other groups; ...(for example) households that use most of their electricity in the evening, and those with steady daytime use; The provider can then design a tariff or an efficiency campaign aimed at each segment; It can also reveal groups the provider had not thought to look for;	
HL 4(b) A4.3.5	Award [2] max Award [1] per point up to [2] max. Association rule discovery finds attributes or events that frequently occur together in the data; ...(for example) households that run a tumble dryer in the evening also tend to run an electric heater at the same time; The provider could use such a rule to predict demand peaks or to target advice about shifting usage to cheaper hours;	2
HL 5(a) A4.3.6	Award [3] max Award [1] per point up to [3] max. Exploitation means choosing the action that has given the best reward so far; Exploration means trying a different action to discover whether something better exists; An agent that only exploits may settle on a merely adequate schedule and never find the best one; An agent that only explores never benefits from what it has already learned; ...(so) the agent must balance the two, typically exploring more early on and exploiting more as its estimates improve;	3
HL 5(b) A4.3.7	Award [2] max Award [1] per correct operator, [2] max. Selection (choosing the fitter routes to become parents); Crossover (combining parts of two parent routes to produce a child); Mutation (randomly altering part of a route to maintain variety); Any two of the three. Do not credit 'fitness function', which evaluates rather than produces the new generation, or 'termination'.	2
HL 6(a) A4.3.8	Award [3] max Award [1] for two or more inputs each with its own weight; [1] for the summation with a bias added; [1] for the activation function producing a single labelled output. The diagram should show inputs x_1, x_2, x_3 entering a node, each along an arrow labelled with a weight w_1, w_2, w_3; Inside the node, the weighted inputs are summed and a bias b is added; That sum is passed through an activation function, which produces the single output y; Accept any conventional layout. All five items named in the question (inputs, weights, bias, activation function, output) must be labelled for full marks. Deduct nothing for the choice of activation function shown.	3
HL 6(b) A4.3.9	Award [2] max Award [1] for the point and [1] for the expansion. Any one reason. A CNN's convolutional layers apply filters across the whole image; ...(so) it can recognise a feature such as a crack wherever it appears in the photograph, whereas an MLP treats each pixel position separately;	2

Q	ANSWER	MARKS
	A CNN shares weights across the image and uses pooling layers to reduce size; ...(so) it needs far fewer parameters than a fully connected MLP on the same image, and trains on less data; A CNN builds a spatial hierarchy, combining edges into shapes and shapes into objects; ...(so) it captures the structure of an image, which an MLP loses when the pixels are flattened into a single list;	

Total for this paper: 34 marks