

THEME A: CONCEPTS · A4.1

Machine learning fundamentals — Mark scheme

33 marks · covers A4.1.1–A4.1.2

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
1(a) A4.1.1	Award [3] max Award [1] per point up to [3] max. Both must be addressed and the contrast made clear. Supervised learning trains on data that has already been labelled with the correct answer, whereas unsupervised learning trains on unlabelled data; In supervised learning the algorithm learns a mapping from inputs to a known output, so it can predict the label of new data; In unsupervised learning the algorithm finds structure or patterns in the data by itself, such as groupings, with no correct answer to compare against; ... (so) the labelled leaf photographs suit supervised learning, while the unlabelled farm records suit unsupervised learning;	3
1(b) A4.1.1	Award [1] max Unsupervised learning; Accept clustering.	1
1(c) A4.1.1	Award [2] max Award [1] per point up to [2] max. Must be applied to the comments. Sentiment analysis classifies each written comment by the opinion it expresses, for example positive, negative or neutral; ... (so) thousands of comments can be summarised automatically instead of being read one by one; The co-operative could track whether opinion is improving or worsening over time; It could flag the negative comments for a person to follow up, or trace complaints back to a particular product or member farm;	2
2(a) A4.1.1	Award [3] max Award [1] per point up to [3] max. This is transfer learning; the candidate need not name it to gain the marks. Start from the public model, keeping the general plant features it has already learned such as edges, shapes and textures; Replace or reset the final output layer so that it produces the two classes the co-operative needs, healthy and diseased; Freeze the earlier layers so their learned features are not lost, and retrain only the final layers;	3

Q	ANSWER	MARKS
	Train those layers on the 400 labelled photographs, then evaluate the result on photographs held back from training; Optionally unfreeze more layers and fine-tune the whole model at a low learning rate;	
2(b) A4.1.1	Award [3] max Award [1] per point up to [3] max. Must be linked to the co-operative's situation. The co-operative has only 400 labelled photographs, which is far too few to train a large model from scratch; A model trained from scratch on 400 images would overfit and perform poorly on new leaves; The publicly available model already recognises general plant features, so only the distinction between healthy and diseased has to be learned; Training is therefore much faster and needs far less computing power; Getting an expert to label many thousands more photographs would be slow and expensive;	3
3(a) A4.1.1	Award [3] max Award [1] per point up to [3] max. Deep learning uses artificial neural networks with many hidden layers, rather than a single layer or a non-network algorithm; The features are learned automatically from the raw data rather than being chosen and engineered by a person; Each layer builds on the one before it, so successive layers represent increasingly complex patterns; It generally needs far more training data and far more processing power than other approaches; ...(which) is why it suits tasks such as image and speech recognition where the useful features are hard to specify by hand;	3
3(b) A4.1.1	Award [2] max Award [1] for the correct type and [1] for the justification. No mark for the type alone. Reinforcement learning; The tractor is not given labelled correct answers; it learns from the reward score it receives after each attempt, adjusting its behaviour to maximise the cumulative reward;	2
4(a) A4.1.2	Award [4] max Award up to [2] per factor (point plus expansion), [4] max, for two different factors. Processing Training a model on 90,000 high-resolution images requires enormous parallel computation; ...(so) hardware with many parallel cores, such as GPUs or TPUs, is needed to finish training in a reasonable time; Storage 90,000 high-resolution photographs occupy a large amount of space, and the data must be read repeatedly during training; ...(so) enough fast storage is needed or the processor will sit idle waiting for data; Scalability The training is repeated each season and the data set will grow; ...(so) the chosen hardware must be able to expand, or be rented in larger amounts, without being replaced; Cost and utilisation Powerful hardware is expensive and would sit unused between seasons;	4

Q	ANSWER	MARKS
	...(so) the co-operative must weigh purchase against renting capacity when needed;	
4(b) A4.1.2	Award [2] max Award [1] for the point and [1] for the expansion. Any one reason. A laptop has few processing cores and limited memory; ...(so) training that a GPU cluster completes in hours could take weeks, or fail to fit in memory at all; A laptop cannot dissipate the heat generated by sustained full-load computation; ...(so) it would throttle its own speed or risk damage over a long training run;	2
5(a) A4.1.2	Award [2] max Award up to [1] per difference, [2] max. A GPU is a general parallel processor originally designed for graphics, whereas a TPU is an application-specific chip designed only for machine learning; A TPU is optimised for the tensor and matrix operations that neural networks use, so it is faster and more energy-efficient for that work; A GPU is more flexible and can be used for other parallel tasks, whereas a TPU is of little use outside machine learning;	2
5(b) A4.1.2	Award [3] max Award [1] per point up to [3] max. The co-operative trains only once a season, so an owned cluster would sit idle most of the year; Cloud platforms are paid for by usage, so the co-operative pays only for the hours it actually trains; There is no large up-front purchase, and no need to house, cool, power or maintain the hardware; Capacity can be scaled up as the data set grows, without buying new machines; The provider keeps the hardware current, so the co-operative is not left with obsolete equipment; Accept a counter-point as part of a balanced answer, for example ongoing cost or the need to upload large data sets.	3
6(a) A4.1.2	Award [3] max Award [1] per point up to [3] max. Must be linked to the field conditions. The fields have no reliable network coverage, so the tractor could not depend on reaching a cloud service; Running the model locally means it keeps working when there is no connection at all; There is no round trip to a distant server, so the decision is made in a fraction of a second; ...(which) matters because the tractor must react to what it sees while it is moving; Only results, rather than continuous high-resolution video, need to be sent back later, saving bandwidth and cost;	3
6(b) A4.1.2	Award [2] max Award [1] for the point and [1] for the expansion. Any one limitation. An edge device has far less processing power and memory than the training hardware; ...(so) the model may have to be simplified or compressed before it will run at an acceptable speed; The device runs from the tractor's limited power supply and must not overheat; ...(so) sustained heavy computation is not possible;	2

Q	ANSWER	MARKS
	Storage on the device is limited; ...(so) it can hold the model and recent data but not the full training set;	

Total for this paper: 33 marks