

Solutions

Each answer shows the steps, then the **result**.

2.1

- (i) **unsupervised**
- (ii) **supervised**
- (iii) **reinforcement**

2.2

- (i) **regression**
- (ii) **classification**

2.3

- $\frac{117}{130} \times 100\%$
- **90.0%**

2.4

- the model learns the training data, including its noise, so well that it fails on new data
- it scores much lower on the unseen test set than on the training set

3.1 A

- inputs with the right answers

3.2 B

- shows how it will do in real use

3.3 B

- bias from unrepresentative training data

3.4

(a)

- **supervised** learning: past students are labelled pass or fail
- **classification**

(b)

- split past records into a training set and a test set
- train the model on the training set
- measure its predictions on the unseen test set before using it

(c)

- predicting “pass” for everyone would score 92% too
- the school needs to know how many of the real failing students it finds

(d) Any two issues, each with a matching action, e.g.:

- bias: postcode may stand for family income, so students from poorer areas are labelled unfairly; remove postcode and check results by group
- privacy: students’ personal data is used without their knowledge; tell students and families, use only needed data, keep it secure
- labelling a student may lower expectations of them; use predictions only to offer extra help, with a teacher deciding