Solutions

Marking: **M1** a correct method, **A1** the correct answer, **B1** an independent mark.

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

2.1

$\bar{x} = 3$, $\bar{y} = 4$

$b = 9/10 = 0.9$

$\hat{y} = \mathbf{0.9x + 1.3}$

2.2

$4 - 4.9 = \mathbf{-0.9}$

$0.9 \times 6 + 1.3 = \mathbf{6.7}$

2.3

$z = -1$

$\sigma(-1) \approx \mathbf{0.27}$: predict fail

$P = 0.5$ when $z = 0$: $x = \mathbf{6}$ hours

2.4

$k = 1$: cross

$k = 3$: circle, 2 votes to 1

3.1

(a) degree 2: it has the lowest validation error

(b) they pass through every training point, noise included, so they swing wildly between points and predict new data badly

(c) more training data; a simpler model, or regularisation that penalises large weights

(d) with enough parameters any data can be fitted exactly, so it says nothing about new data

3.2

(a) P1 $(1, 1)$: 4; P2 $(2, 4)$: 2; P3 $(4, 4)$: 2; P4 $(5, 1)$: 4; P5 $(3, 6)$: 3

(b) the 3 nearest: P2 (A), P3 (B), P5 (B); class **B**

(c) the feature with large numbers dominates the distance, so scale the features first

3.3

(a) classification (pass or fail); logistic regression

(b) its weights show which features matter, so teachers can understand it; it gives a probability, and is fast with no need to store all 600 students

(c) missing a student who then fails costs more than giving help to one who did not need it, so a lower threshold catches more at-risk students

(d) as a list for teachers to review, not an automatic label; check the results each term, and tell students only about the extra help, not the score

exercises.lol exercises.com.cn