Solutions

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

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

2.1

$w_1 = 0.6$

$w_2 = 1.08$

$w_3 = 1.08 + 0.2 \times 1.92 = \mathbf{1.464}$

2.2

$w \to 6 \to 0 \to 6$: each step jumps to the mirror point, so it never settles

$w - 3$ is multiplied by $1 - 2\eta$ each step, so it converges when $0 < \eta < 1$

2.3

$50\,000/100 = \mathbf{500}$ steps per epoch

10 000 steps for 20 epochs

2.4

$e^{-2/10} \approx \mathbf{0.82}$

$e^{-2} \approx \mathbf{0.14}$

at high T the search moves freely; as T falls it settles into a minimum

3.1

(a) hidden: $784 \times 128 + 128 = 100\,480$; output: $128 \times 10 + 10 = 1290$; total **101 770**

(b) $\lceil 60\,000/64 \rceil = 938$ per epoch; **9380** steps

(c) to choose settings such as η and when to stop on data the weights never trained on, so overfitting shows up

3.2

(a) slow, steady fall: too small; fast fall that levels off low: good; rising, jumpy curve: too large

(b) start with a large η and reduce it during training: fast progress early, and small steps near the minimum so it settles

3.3

(a) every swap of two lessons makes the clashes the same or worse: it is a local minimum, and hill climbing never accepts a worse move

(b) random restarts from new timetables; simulated annealing, which sometimes accepts worse moves

(c) $5 \times 0.95^n < 0.1$: $n > \ln(0.02)/\ln(0.95) \approx 76.3$, so **77** rounds

exercises.lol exercises.com.cn

(d) simulated annealing with several restarts, keeping the best result; add hard rules (no teacher in two rooms) as large costs, and check the final timetable by hand