

Exercise walk-through

13.4

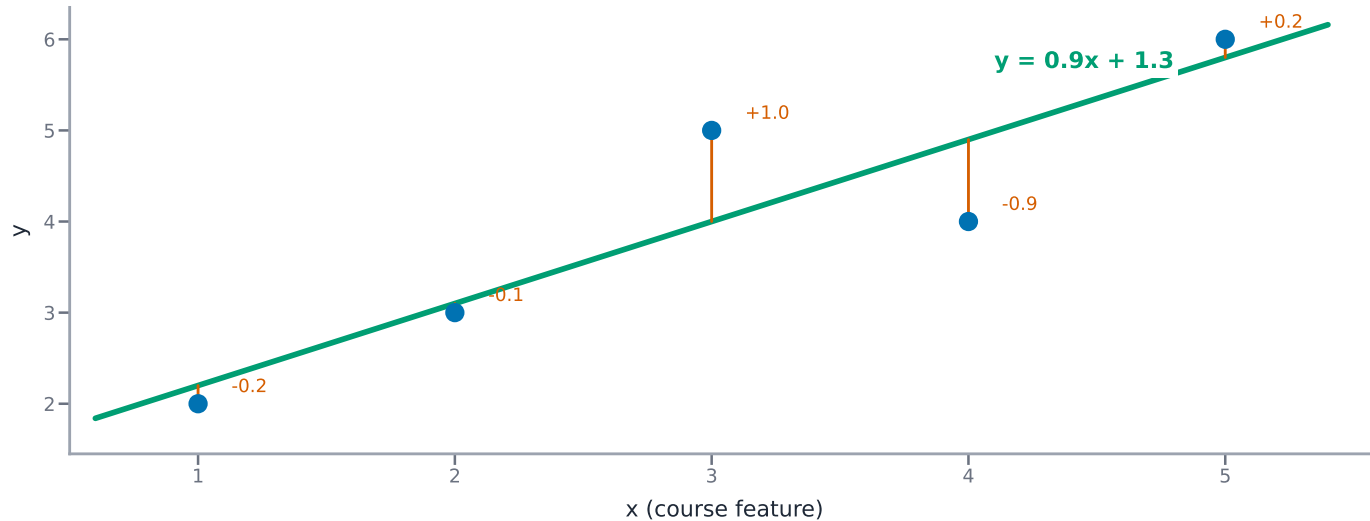
Regression and classification

University Computer Science

Regression and classification: the boundary and its bill

- 1 **Line** Points (1,2)(2,3)(3,5)(4,4)(5,6): $w = 9/10$, $b = 1.3$, residuals -0.2 -0.1 1.0 -0.9 0.2, SSE 1.9.
- 2 **Boundary** w_1 2, w_2 -1, b -1: the line $x_2 = 2x_1 - 1$; (2,2) scores +1 late, (2,5) scores -2 on time; margin = score magnitude.
- 3 **Overfit** 4th-degree poly: SSE 0 on train, about 42.7 held-out vs the line 0.55 - model choice by held-out error.

Decision boundary and data



residuals -0.2, -0.1, +1.0, -0.9, +0.2 — SSE 1.9; a 4th-degree poly fits 0 here but held-out 42.7 vs the line 0.55

The fitted line with residuals; the boundary where the answer flips.

Exercises 2.1-2.3

2.1 What does the logistic boundary $2x_1 - x_2 - 1 = 0$ predict at $(2, 2)$? [3]

2.2 Where does the margin's doubt live? [3]

2.3 Training error or held-out error – which is the product? [3]

Exercises 2.1·2.3

2.1 What does the logistic boundary $2x_1 - x_2 - 1 = 0$ predict at $(2, 2)$? [3]

2.1

- Score $2(2) - 2 - 1 = 1 > 0$: predicted late ✓.

2.2 Where does the margin's doubt live? [3]

2.2

- Points nearest the boundary, where scores are small ✓.

2.3 Training error or held-out error – which is the product? [3]

2.3

- Held-out error: training error is an advertisement ✓.

Exercise 3.1

3.1 Fit the least-squares line to $(1, 3), (2, 3), (3, 4), (4, 6), (5, 6)$.

Compute $\bar{x}$, $\bar{y}$, the slope numerator and denominator, and w, b . Give the residuals and the SSE. Predict at $x = 6$ and state one reason the prediction could mislead. [8]

Exercise 3.1

3.1 Fit the least-squares line to $(1, 3), (2, 3), (3, 4), (4, 6), (5, 6)$.

Compute $\bar{x}$, $\bar{y}$, the slope numerator and denominator, and w, b . Give the residuals and the SSE. Predict at $x = 6$ and state one reason the prediction could mislead. [8]

3.1

- $\bar{x} = 3, \bar{y} = 4.4$ ✓.
- Deviations from the means give numerator 9.0 and denominator 10 ✓.
- Slope $w = 9/10 = 0.9$; intercept $b = 1.7$ ✓.
- Residuals 0.4, $-0.5, -0.4, 0.7, -0.2$; SSE 1.1 ✓.
- Prediction at 6: $0.9 \times 6 + 1.7 = 7.1$ – an extrapolation beyond the data range ✓.

Exercise 3.2

3.2 A logistic model has $w_1 = 1, w_2 = 2, b = -4$ over features (distance in km, traffic index).

Write the boundary equation and solve it for the traffic axis. Classify the points $(2, 1)$ and $(2, 1.5)$ with their scores. State which point the model is more confident about and why. Explain the design consequence of an unscaled traffic index on the fit. [7]

Exercise 3.2

3.2 A logistic model has $w_1 = 1, w_2 = 2, b = -4$ over features (distance in km, traffic index).

Write the boundary equation and solve it for the traffic axis. Classify the points $(2, 1)$ and $(2, 1.5)$ with their scores. State which point the model is more confident about and why. Explain the design consequence of an unscaled traffic index on the fit. [7]

3.2

- Boundary: $x_1 + 2x_2 - 4 = 0$, i.e. $x_2 = (4 - x_1)/2$ ✓.
- $(2, 1)$: score $2 + 2 - 4 = 0$ – exactly on the boundary, maximally doubtful ✓.
- $(2, 1.5)$: score $2 + 3 - 4 = +1 > 0$: positive class, with margin 1 ✓.
- More confident at $(2, 1.5)$: its score magnitude is larger – distance from the boundary ✓.
- Unscaled index: its weight absorbs raw magnitude, the loss surface ravines (unit 13.5) – standardise before fitting ✓.