UNIVERSITY COMPUTER SCIENCE • EXERCISE SHEET

13.4 Regression and classification

Name: _____ Class: _____ Date: _____ **Total: 24 marks**

KEY WORDS linear regression • decision boundary

Objective

Fit by least squares, read a boundary, and price complexity in held-out error.

1 Key knowledge and worked examples

■ **The fit**

Course points $(1,2),(2,3),(3,5),(4,4),(5,6)$: the line $\hat{y} = 0.9x + 1.3$ with SSE 1.9 ✓.

■ **Worked example: the overfit**

The 4th-degree polynomial scores SSE 0 on those five points but ≈ 42.7 on the held-out five; the line scores 0.55 ✓.

2 Practice

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]

3 Exam-style questions

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]

3.2 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]