



Preprocessing: Scaling & Encoding

SCORE

___ / 20

Chapter: AI & Machine Learning · Level ★★★ ·

Time: 30 min

Name _____ Class _____ Date _____

AFTER THIS WORKSHEET YOU CAN

✓ StandardScaler and MinMax ✓ Why scaling matters ✓ One-hot encoding ✓ Handle unseen categories



Section A · Concept Check

• BEGINNER

4 × 1 = 4

- 1 StandardScaler outputs features with:
☐ A. mean 0, std 1 ☐ B. 0..255
☐ C. only positives ☐ D. integers
- 2 MinMaxScaler maps to:
☐ A. [0,1] by default ☐ B. [-100,100]
☐ C. z-scores ☐ D. ranks
- 3 Categorical text becomes numbers via:
☐ A. one-hot encoding ☐ B. rounding
☐ C. hashing only ☐ D. deletion
- 4 Distance-based models (KNN/SVM) need:
☐ A. scaled features ☐ B. raw pixels
☐ C. unscaled data ☐ D. strings



Section B · Problem Solving

• INTERMEDIATE

2 + 3×3 = 11

- 5 Standardisation: $(x - \text{mean}) / \text{_____}$
- 6 One-hot creates one _____ column per category.
- 7 Why does KNN fail with unscaled salary + age?

- 8 One-hot encode colors red/green/blue – result shape?

- 9 Standard vs MinMax – when which?



Section C · Challenge

• CHALLENGE

5

- 10 What is data leakage in scaling?

🧠 Reflection — the most useful thing I learned: _____

A ___/4 B ___/11 C ___/5 Total ___/20

Teacher's Signature _____

Parent's Signature _____

✂️ ANSWER KEY – FOLD OR CUT BEFORE HANDING OUT

1-A 2-A 3-A 4-A | 5 std 6 binary/indicator 7 = Salary's huge range dominates the distance metric
8 = 3 binary columns 9 = Standard for normal-ish data/linear models; MinMax for bounded inputs
| 10 = Fitting the scaler with test data - inflates evaluation

