After Full Marks: Did the Model Really Learn?

Plan 35-45 minutes: predict on paper, then check on one projected computer. The local-model lesson needs a computer to run its download. Split advanced calculations across two sessions.

Learning goal

Count missed cats and false cat alarms separately, and reserve test data before tuning.

Before starting

Complete the perceptron and training-data lessons; use fractions and percentages.

Materials

The new worksheet, ten labels (one cat, nine non-cats) and training, validation and test envelopes.

Class sequence (5 / 10 / 10 / 10 min, plus 5-10 min discussion)

1. Predict: a model always says not-cat. Score it on one cat and nine non-cats, then score just the cat.

2. Try: six truths are cat,cat,no,no,no,no; predictions are cat,no,cat,no,no,no. Assign every example to TP, FN, FP or TN.

3. Observe: split into three envelopes first. Train weights on training data, choose thresholds/models with validation data and open final test data after selection.

4. Challenge: put copies of one original picture in training and test. Explain why this is not an independent new problem, then split by source.

After Full Marks: Did the Model Really Learn?

Worked example

Always not-cat gives 9/10=90% accuracy but cat recall 0/1=0%. The second set has TP=1,FN=1,FP=1,TN=3 and accuracy 4/6.

Check understanding: ask before showing an answer

After repeatedly tuning a threshold against test results, is that still an independent final exam?

Expected answers and worksheet check

No; test information has influenced selection, so a new independent test is needed. Second set: precision=1/(1+1)=50%, recall=1/(1+1)=50%, accuracy=4/6≈66.7%. This tiny exercise is not a deployment-performance estimate. Challenge: nine cats, one non-cat, always cat gives TP=9,FP=1,FN=0,TN=0. Non-cat recall is 0/1: every non-cat is a false alarm. Keep duplicate/source-related images together in one split. Choose thresholds with validation data.

Keep the claim within its limits

Accuracy, precision and recall answer different questions. No single metric suits every use; a zero denominator must be marked undefined.

Adapt the lesson

For support, work through just the example, read prompts aloud and use counters or paper grids. For extension, change one condition at a time, write a prediction and test it. When answers differ, inspect rules, encoding and data before judging a child’s answer.

At home

Show a parent the 90% score, then reveal how many cats there were and ask what else should be counted.

Continue learning

Neural Networks with Layers / CNN: Find a Line with One Shared Ruler / Your Own Tiny Language Model: Count, Then Generate

Teacher reference (links also on the lesson page)

scikit-learn: Common pitfalls: https://scikit-learn.org/stable/common_pitfalls.html