Unit 10 · How Machines Learn

WORKSHEET

Full marks 30

Name _____ Class _____ Roll no. ____ Date _____ Marks _____ / 30

1 Key words [4]

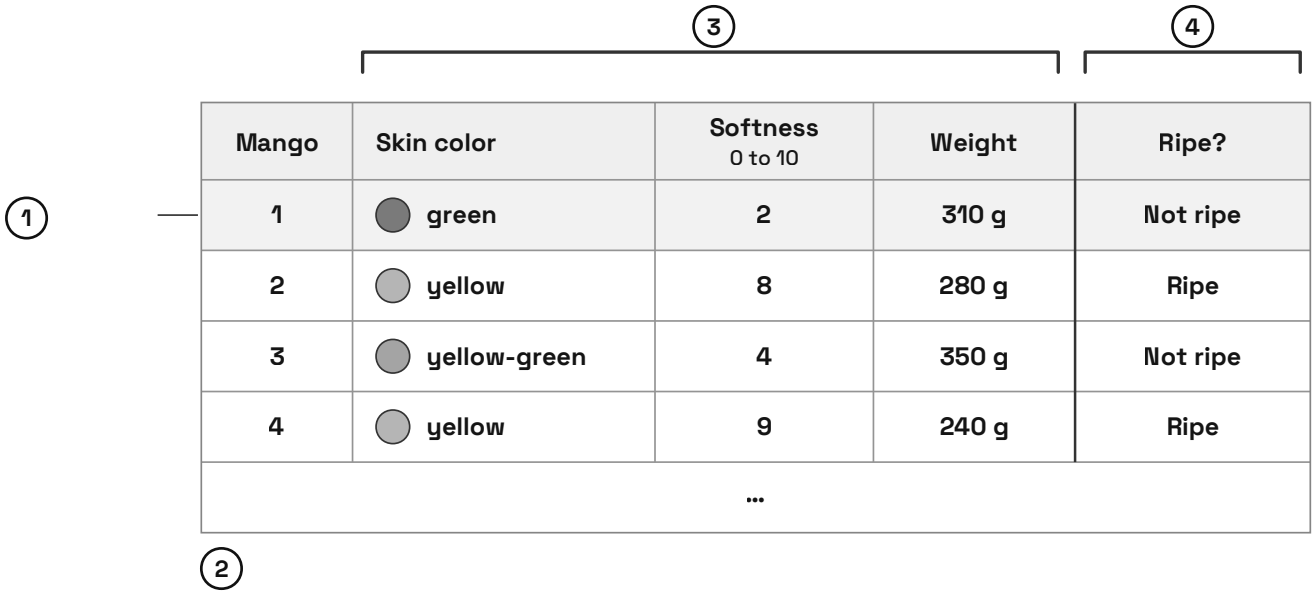
Fill in the blanks with words from the box. Two words in the box are not needed. Each blank carries 0.5 mark.

accuracy	**biased**	**dataset**	**filter bubble**	**labeling**	**nonexamples**	**one-sided data**	**pixel**	**test set**	**training set**

A collection of examples is a _____. Before training, the examples are split in two: the model learns from the _____, and the _____ is kept apart and used only to check the model at the end. Giving every example its right answer is called _____; it takes a long time, and people make mistakes. Things that are not what the model is looking for, shown to it so that it learns where its answer ends, are _____. The share of the test examples a model gets right is its _____. A model that works worse for one group than for others, because of how it was made, is _____. A common cause is _____: too few examples of some of the things the model will meet.

2 Label the diagram [2]

The diagram shows training data for a mango model, as a table. Write what belongs in place of 1 to 4. Each carries 0.5 mark.



Mango	Skin color	Softness 0 to 10	Weight	Ripe?
1	green	2	310 g	Not ripe
2	yellow	8	280 g	Ripe
3	yellow-green	4	350 g	Not ripe
4	yellow	9	240 g	Ripe
…				

1. _____ 2. _____

3. _____ 4. _____

3 Multiple choice [5]

Choose the correct option for each of the following questions.

i. In a dataset for a model that predicts whether a bus from Kathmandu to Pokhara will arrive late, which is the **label**? [1]

a. the time it left Kathmandu
b. the number of passengers
c. whether that bus did arrive late
d. whether it was raining

ii. A model gets 100% on its training examples and 60% on its test set. What does this show? [1]

a. It is an excellent model.
b. The test set is wrong.
c. It remembers its training examples but does much less well on new ones; 60% is the honest measure.
d. It needs a faster computer.

iii. A model was tested on 60 photos and got 51 right. What is its accuracy? [1]

a. 51%
b. 60%
c. 85%
d. 90%

iv. A team's model does badly on Bambaiya mangoes. What will help most? [1]

a. adding many more examples of the kinds it already gets right
b. testing it on fewer Bambaiya
c. adding examples of ripe and unripe Bambaiya to the training data
d. making the model run faster

v. Which of these is a decision made while **testing** that can give a misleading result? [1]

a. giving every example a label
b. reporting only the total accuracy, not the accuracy for each group
c. keeping the test set apart from the training set
d. choosing test examples at random from all the data

4 Calculations [6]

Show your working.

a. Model A got 42 of 50 test photos right. Model B got 70 of 80 test photos right. Which model is more accurate? [2]

Answer: _____

b. A crop disease model was tested on 120 leaf photos and got 102 right. **[4]**

i. What is its accuracy?

ii. Of the photos, 90 were of rice leaves and 30 of maize leaves. It got 87 rice photos right and 15 maize photos right. Find its accuracy for each crop.

iii. What do your answers show?

Answer: _____

5 Short answers **[7]**

Write short answers to the following questions.

a. A team in Pokhara wants a model that says each morning whether it will rain in the afternoon. For every day of the past five years they have: the morning's humidity, the cloud cover at 9 a.m., the wind speed, the month, and whether it rained that afternoon. Name the features and the label. Suggest one more feature that might help, and one that would not. **[4]**

b. What are nonexamples? Give two nonexamples that would be useful for a model that finds momos in photos, and say why look-alikes are the most useful. **[3]**

6 Long answer [4]

Write a long answer to the following question.

What is bias in an AI model? Explain, with an example, how one-sided training data makes a biased model.

7 Think [2]

Think, then answer.

A school wants a model to choose who gets a scholarship. Its training data is the last ten years of students, with their marks, their village, and whether they finished Grade 12. The model learns that students from two villages rarely finished. Should it give those villages' students lower scores? Explain.

8 Activity: Does voice typing hear everyone? (not marked)

Work in a group, with a teacher's phone that has voice typing in Nepali. Choose one Nepali sentence of about 10 words. Each person in the group reads it aloud to the voice typing. Count the words it typed right, and work out its accuracy for each speaker.

Speaker (where they grew up)	Words spoken	Words typed right	Accuracy