← **Back**

Bird Recognition in the City of Peacetopia (Case Study)

Graded Quiz • 1h 15m

English

Due Nov 19, 11:59 PM WIB

⚠ Try again once you are ready

Grade received 66.66%

Latest Submission Grade 66.67%

To pass 80% or higher

Try again

1. **Problem Statement**

1 / 1 point

This example is adapted from a real production application, but with details disguised to protect confidentiality.

You are a famous researcher in the City of Peacetopia. The people of Peacetopia have a common characteristic: they are afraid of birds. To save them, you have **to build an algorithm that will detect any bird flying over Peacetopia** and alert the population.

The City Council gives you a dataset of 10,000,000 images of the sky above Peacetopia, taken from the city's security cameras. They are labeled:

- y = 0: There is no bird on the image
- y = 1: There is a bird on the image

Your goal is to build an algorithm able to classify new images taken by security cameras from Peacetopia.

There are a lot of decisions to make:

- What is the evaluation metric?
- How do you structure your data into train/dev/test sets?

Metric of success

The City Council tells you that they want an algorithm that

1. Has high accuracy.
2. Runs quickly and takes only a short time to classify a new image.
3. Can fit in a small amount of memory, so that it can run in a small processor that the city will attach to many different security cameras.

Note: Having three evaluation metrics makes it harder for you to quickly choose between two different algorithms, and will slow down the speed with which your team can iterate. True/False?

● True

○ False

Expand

✓ Correct

2. After further discussions, the city narrows down its criteria to:

1 / 1 point

- "We **need** an algorithm that can let us know a bird is flying over Peacetopia as accurately as possible."
- "We *want* the trained model to take no more than 10 sec to classify a new image."

- "We *want* the model to fit in 10MB of memory."

If you had the three following models, which one would you choose?

○

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

●

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

○

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

○

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

Expand

✓ **Correct**
Correct! This model has the highest test accuracy, the prominent criteria you are looking for, compared with other models, and also has a runtime <10 seconds and memory size < 10MB.

3. Which of the following best answers why it is important to identify optimizing and satisficing metrics? **1 / 1 point**

● Identifying the metric types sets thresholds for satisficing metrics. This provides explicit evaluation criteria.

○ Identifying the optimizing metric informs the team which models they should try first.

○ It isn't. All metrics must be met for the model to be acceptable.

○ Knowing the metrics provides input for efficient project planning.

Expand

✓ **Correct**
Yes. Thresholds are essential for evaluation of key use case constraints.

4. You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal? **0 / 1 point**

● The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.

○ The emphasis on the training set will allow us to iterate faster.

○ With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.

○ The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.

Expand

⊗ **Incorrect**
No. There is not enough information to consider the satisficing metrics yet.

5. Now that you've set up your train/dev/test sets, the City Council comes across another 1,000,000 images from social media and offers them to you. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. Which of the following is the best use of that additional data? **0 / 1 point**

● Do not use the data. It will change the distribution of any set it is added to.

○ Add it to the dev set to evaluate how well the model generalizes across a broader set

○ Add it to the dev set to evaluate how well the model generalizes across a broader set.

○ Add it to the training set.

○ Split it among train/dev/test equally.

Expand

⊗ **Incorrect**
No. The data can contribute to training the model.

6. One member of the City Council knows a little about machine learning and thinks you should add the 1,000,000 citizens' data images proportionately to the train/dev/test sets. You object because:

0 / 1 point

○ If we add the images to the test set then it won't reflect the distribution of data expected in production.

○ The training set will not be as accurate because of the different distributions.

● The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.

○ The additional data would significantly slow down training time.

Expand

⊗ **Incorrect**
No. The important issue is mixing distributions.

7. Human performance for identifying birds is < 1%, training set error is 5.2% and dev set error is 7.3%. Which of the options below is the best next step?

0 / 1 point

● Validate the human data set with a sample of your data to ensure the images are of sufficient quality.

○ Try an ensemble model to reduce bias and variance.

○ Train a bigger network to drive down the >4.0% training error.

○ Get more data or apply regularization to reduce variance.

Expand

⊗ **Incorrect**
No. Unless you have strong reasons to believe the labeling is suspect, it's back to the drawing board.

8. You ask a few people to label the dataset so as to find out what is human-level performance. You find the following levels of accuracy:

1 / 1 point

Bird watching expert #1	0.3% error
Bird watching expert #2	0.5% error
Normal person #1 (not a bird watching expert)	1.0% error
Normal person #2 (not a bird watching expert)	1.2% error

If your goal is to have "human-level performance" be a proxy (or estimate) for Bayes error, how would you define "human-level performance"?

○ 0.0% (because it is impossible to do better than this)

○ 0.75% (average of all four numbers above)

● 0.3% (accuracy of expert #1)

○ 0.4% (average of 0.3 and 0.5)

Expand

✓ Correct

9. A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error. True/False?

1 / 1 point

○ False.

● True.

Expand

✓ Correct

Yes. By definition, human level error is worse than Bayes error.

10. Which of the following best expresses how to evaluate the next steps in your project when your results for human-level performance, train, and dev set error are 0.1%, 2.0%, and 2.1% respectively?

1 / 1 point

● Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.

○ Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.

○ Port the code to the target devices to evaluate if your model meets or exceeds the satisficing metrics.

○ Evaluate the test set to determine the magnitude of the variance.

Expand

✓ Correct

Yes. Always choose the area with the biggest opportunity for improvement.

11. You also evaluate your model on the test set, and find the following:

1 / 1 point

Human-level performance	0.1%
Training set error	2.0%
Dev set error	2.1%
Test set error	7.0%

What does this mean? (Check the two best options.)

☑ You should try to get a bigger dev set.

✓ Correct

☐ You should get a bigger test set.

☑ You have overfit to the dev set.

✓ Correct

☐ You have underfitted to the dev set.

Expand

✓ Correct

Great, you got all the right

Great, you got all the right answers.

12. After working on this project for a year, you finally achieve: Human-level performance, 0.10%, Training set error, 0.05%, Dev set error, 0.05%. Which of the following are true? (Check all that apply.)

1 / 1 point

□ This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.

☑ All or almost all of the avoidable bias has been accounted for.

✓ **Correct**
Yes. Exceeding human performance makes the identification of avoidable bias very challenging.

☑ You are close to Bayes error and possible overfitting.

✓ **Correct**
Yes. By definition, Bayes error cannot be exceeded except for overfitting.

□ With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%

Expand

✓ **Correct**
Great, you got all the right answers.

13. It turns out Peacetopia has hired one of your competitors to build a system as well. Your system and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! However, when Peacetopia tries out your and your competitor's systems, they conclude they actually like your competitor's system better, because even though you have higher overall accuracy, you have more false negatives (failing to raise an alarm when a bird is in the air). What should you do?

1 / 1 point

○ Pick false negative rate as the new metric, and use this new metric to drive all further development.

○ Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.

● Rethink the appropriate metric for this task, and ask your team to tune to the new metric.

○ Ask your team to take into account both accuracy and false negative rate during development.

Expand

✓ **Correct**

14. You've handily beaten your competitor, and your system is now deployed in Peacetopia and is protecting the citizens from birds! But over the last few months, a new species of bird has been slowly migrating into the area, so the performance of your system slowly degrades because your model is being tested on a new type of data. There are only 1,000 images of the new species. The city expects a better system from you within the next 3 months. Which of these should you do first?

0 / 1 point

○ Put them into the dev set to evaluate the bias and re-tune.

○ Augment your data to increase the images of the new bird.

● Add the new images and split them among train/dev/test.

○ Add hidden layers to further refine feature development.

Expand

✗ **Incorrect**
No. The number of new images is too small to make a difference.

15. The City Council thinks that having more Cats in the city would help scare off birds. They are so happy with your work on the Bird detector that they also hire you to build a Cat detector. (Wow Cat detectors are just incredibly useful, aren't they?) Because of years of working on Cat detectors, you have such a huge dataset of 100,000,000 cat images that training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.)

1 / 1 point

□ Having built a good Bird detector, you should be able to take the same model and hyperparameters and just apply it to the Cat dataset, so there is no need to iterate.

☑ Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.

✓ **Correct**

☑ If 100,000,000 examples is enough to build a good enough Cat detector, you might be better off training with just 10,000,000 examples to gain a $\approx$10x improvement in how quickly you can run experiments, even if each model performs a bit worse because it's trained on less data.

✓ **Correct**

☑ Needing two weeks to train will limit the speed at which you can iterate.

✓ **Correct**

Expand

✓ **Correct**

Great, you got all the right answers.