Bird Recognition in the City of Peacetopia (Case Study)

Congratulations! You passed!

Grade received 100%	Latest Submission Grade 100%	To pass 80% or higher

Go to next item

1. This example is adapted from a real production application, but with details disguised to protect confidentiality. **1 / 1 point**

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 the following 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.

You meet with them and ask for just one evaluation metric. True/False?

- (●) True:
- ( ) False

Timed quizzes

Expand

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

2. The city asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria? **1 / 1 point**

- ○ Suggest to them that they focus on whichever criterion is to be optimized and then eliminate the other two.
- ○ Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.
- ◉ Suggest to them that they define which criterion is to be optimized. Then, set thresholds for the other two.

Expand

Correct
Yes. The thresholds provide a way to evaluate models head to head.

3. The essential difference between an optimizing metric and satisficing metrics is the priority assigned by the stakeholders. True/False? **1 / 1 point**

- ◉ False
- ○ True

Expand

Correct
Yes. Satisficing metrics have thresholds for measurement and an optimizing metric is unbounded.

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? **1 / 1 point**

- (●) 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.
- ( ) 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.

Expand

Correct
Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

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? **1 / 1 point**

- ( ) Do not use the data. It will change the distribution of any set it is added to.
- ( ) Split it among train/dev/test equally.
- (●) Add it to the training set.
- ( ) Add it to the dev set to evaluate how well the model generalizes across a broader set.

Expand

Correct
Yes. It is not a problem to have different training and dev distributions. Different dev and test distributions would be an issue.

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: **1 / 1 point**

- The additional data would significantly slow down training time.
- The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- The training set will not be as accurate because of the different distributions.
- If we add the images to the test set then it won't reflect the distribution of data expected in production.

Expand

Correct
Yes. Using the data in the training set could be beneficial, but you wouldn't want to include such images in your test set as they are not from the expected distribution of data you'll see in production.

7. You train a system, and its errors are as follows (error = 100%-Accuracy): **1 / 1 point**

Training set error	4.0%
Dev set error	4.5%

This suggests that one good avenue for improving performance is to train a bigger network so as to drive down the 4.0% training error. Do you agree?

- ○ No, because this shows your variance is higher than your bias.
- ◉ No, because there is insufficient information to tell.
- ○ Yes, because this shows your bias is higher than your variance.
- ○ Yes, because having a 4.0% training error shows you have a high bias.

Expand

✓ Correct

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? **1 / 1 point**

- ○ The average performance of all their ornithologists (0.5%).
- ○ The average of regular citizens of Peacetopia (1.2%).
- ◉ The performance of their best ornithologist (0.3%).
- ○ The average of all the numbers above (0.66%).

Expand

Correct
Yes. The best human performance is closest to Bayes' error.

9. Which of the below shows the optimal order of accuracy from worst to best? **1 / 1 point**

- Human-level performance -> Bayes error -> the learning algorithm's performance.
- Human-level performance -> the learning algorithm's performance -> Bayes error.
- The learning algorithm's performance -> human-level performance -> Bayes error.
- The learning algorithm's performance -> Bayes error -> human-level performance.

Expand

Correct
Yes. A learning algorithm's performance can be better than human-level performance but it can never be better 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**

- Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.
- Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.
- Evaluate the test set to determine the magnitude of the variance.

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

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.)

- [x] You should try to get a bigger dev set.

Correct

- [x] You have overfit to the dev set.

Correct

- [ ] You should get a bigger test set.
- [ ] You have underfitted to the dev set.

Correct
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

- [x] 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.

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

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**

- ○ Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- ○ Ask your team to take into account both accuracy and false negative rate during development.
- ◉ Rethink the appropriate metric for this task, and ask your team to tune to the new metric.
- ○ Pick false negative rate as the new metric, and use this new metric to drive all further development.

Expand

Correct

14. 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 data 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? **1 / 1 point**

- ○ Split them between dev and test and re-tune.
- ○ Add pooling layers to downsample features to accommodate the new species.
- ◉ Augment your data to increase the images of the new bird.

- ◯ Put the new species' images in training data to learn their features.

Expand

✓ **Correct**
Yes. A sufficient number of images is necessary to account for the new species.

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. You have a huge dataset of 100,000,000 cat images. Training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.) **1 / 1 point**

- ☐ With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.
- ☑ You could consider a tradeoff where you use a subset of the cat data to find reasonable performance with reasonable iteration pacing.

✓ **Correct**
Yes. This is similar to satisficing metrics where "good enough" determines the size of the data.

- ☑ Accuracy should exceed the City Council's requirements but the project may take as long as the bird detector because of the two week training/iteration time.

✓ **Correct**
Yes. The 10x size increase adds a small amount of accuracy but takes too much time.

- ☑ Given a significant budget for cloud GPUs, you could mitigate the training time.

✓ **Correct**
Yes. More resources will allow you to iterate faster.

Expand

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