

Try again once you are ready

Grade
received **66.66%**

Latest Submission
Grade 66.67%

To pass 80% or
higher

Try again

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

[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 important and then eliminate the

other two.

- ☒ Suggest to them that they define which criterion is most important. Then, set thresholds for the other two.
- ☐ Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.

 Expand

 **Correct**

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

3. Based on the city's requests, which of the following would you say is true?

1 / 1 point

- ☐ Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.
- ☒ Accuracy is an optimizing metric; running time and memory size are satisfying metrics.
- ☐ Accuracy, running time and memory size are all satisfying metrics because you have to do sufficiently well on all three for your system to be acceptable.
- ☐ Accuracy is a satisfying metric; running time and memory size are an optimizing metric.

 Expand

 **Correct**

4. With 10,000,000 data points, what is the best option for train/dev/test splits?

1 / 1 point

- ☐ train - 60%, dev - 30%, test - 10%
- ☐ train - 33.3%, dev - 33.3%, test - 33.3%
- ☐ train - 60%, dev - 10%, test - 30%
- ☒ train - 95%, dev - 2.5%, test - 2.5%

 Expand

 **Correct**

Yes. The size of the data set allows for bias and variance evaluation with smaller data sets.

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.
- ☐ Split it among train/dev/test equally.
- ☒ Add it to the dev set to evaluate how well the model generalizes across a broader set.
- ☐ Add it to the training set.

 Expand

 **Incorrect**

No. Different dev and test distributions would be an issue in evaluating variance because of the different distributions.

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 to the test set. You object because:

1 / 1 point

- ☐ A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.
- ☒ The test set no longer reflects the distribution of data (security cameras) you most care about.

✓ Correct

- ☐ The 1,000,000 citizens' data images do not have a consistent $x \rightarrow y$ mapping as the rest of the data.
- ☒ This would cause the dev and test set distributions to become different. This is a bad idea because you're not aiming where you want to hit.

✓ Correct

↗ Expand

✓ Correct
Great, you got all the right answers.

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

- ☒ Get more data or apply regularization to reduce variance.
- ☐ Try an ensemble model to reduce bias and variance.
- ☐ Train a bigger network to drive down the $>4.0\%$ training error.
- ☐ Validate the human data set with a sample of your data to ensure the images are of sufficient quality.

↗ Expand

✗ Incorrect
No. We do not have test set accuracy.

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

1 / 1 point

- ☒ The best performance of a specialist (ornithologist) or possibly a group of specialists.
- ☐ The performance of the average citizen of Peacetopia.
- ☐ The performance of their volunteer amateur ornithologists.
- ☐ The performance of the head of the City Council.

↗ Expand

✓ Correct
Yes. This is the peak of human performance in this task.


9. Which of the below shows the optimal order of accuracy from worst to best?

0 / 1 point

- ☐ The learning algorithm's performance \rightarrow Bayes error \rightarrow human-level performance.
- ☐ The learning algorithm's performance \rightarrow human-level performance \rightarrow Bayes error.

- ☐ Human-level performance -> the learning algorithm's performance -> Bayes error.
- ☒ Human-level performance -> Bayes error -> the learning algorithm's performance.

 Expand


 **Incorrect**
No. 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.
- ☐ Port the code to the target devices to evaluate if your model meets or exceeds the satisfying metrics.
- ☒ 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.


 Expand

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


11. You've now also run your model on the test set and find that it is a 7.0% error compared to a 2.1% error for the dev set. What should you do? (Choose all that apply)

0 / 1 point


- ☒ Increase the size of the dev set.

 **Correct**
Yes. The dev set performance versus the test set indicates it is overfitting.

- ☒ Try decreasing regularization for better generalization with the dev set.


 **This should not be selected**
No. Decreasing regularization will increase overfitting.

- ☒ Try increasing regularization to reduce overfitting to the dev set.

 **Correct**
Yes. The dev set performance versus the test set indicates it is overfitting.

- ☐ Get a bigger test set to increase its accuracy.

 Expand

 **Incorrect**
You chose the extra incorrect answers.

12. After working on this project for a year, you finally achieve:

1 / 1 point

Human-level performance	0.10%
Training set error	0.05%
Dev set error	0.05%

What can you conclude? (Check all that apply.)

- ☒ It is now harder to measure avoidable bias, thus progress will be slower going forward.

 **Correct**

- ☐ This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- ☒ If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

✓ Correct

- ☐ 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. You and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! Still, when Peacetopia tries out both systems, they conclude they 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?

0 / 1 point

- ☐ Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.
- ☒ Ask your team to take into account both accuracy and false negative rate during development.
- ☐ Apply regularization to minimize the false negative rate.
- ☐ Pick false negative rate as the new metric, and use this new metric to drive all further development.

↗ Expand

✗ Incorrect
No. This is using two optimizing metrics instead of one.

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

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

- ☒ 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.

- ☒ 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.

☐ With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.

☒ 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.