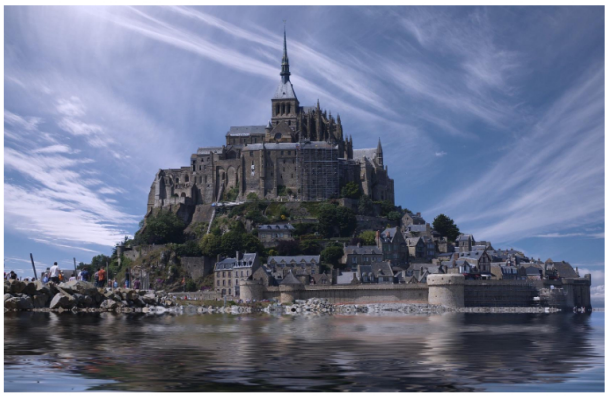
Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | 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?

- ○ False
- ◉ True:

Expand

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

2. After further discussions, the city narrows down its criteria to: 0 / 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
99%	13 sec	9MB

○

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

○

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

○

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

Expand

Incorrect
The runtime doesn't satisfy the requirement from the City Council (it's >10sec).

3. Based on the city's requests, which of the following would you say is true? 1 / 1 point

- (•) Accuracy is an optimizing metric; running time and memory size are satisfying metrics.
- ( ) Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.
- ( ) 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. **Structuring your data** 1 / 1 point

Before implementing your algorithm, you need to split your data into train/dev/test sets. Which of these do you think is the best choice?

( )

Train	Dev	Test
3,333,334	3,333,334	3,333,334

( )

Train	Dev	Test
6,000,000	1,000,000	3,000,000

( )

Train	Dev	Test
6,000,000	3,000,000	1,000,000

(•)

Train	Dev	Test
9,500,000	250,000	250,000

Expand

Correct
Yes.

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

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

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 to the dev set. You object because: (Choose all that apply) 1 / 1 point

- [x] The dev set no longer reflects the distribution of data (security cameras) you most care about.

  ✓ Correct
  Yes. The performance of the model should be evaluated on the same distribution of images it will see in production.

- [ ] The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- [ ] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.
- [x] 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
Yes. Adding a different distribution to the dev set will skew bias.

Expand

Correct
Great, you got all the right answers.

7. You train a system, and the train/dev set errors are 3.5% and 4.0% respectively. You decide to try regularization to close the train/dev accuracy gap. Do you agree? 1 / 1 point

- No, because this shows your variance is higher than your bias.
- Yes, because this shows your bias is higher than your variance.
- No, because you do not know what the human performance level is.
- Yes, because having a 4.0% training error shows you have a high bias.

Expand

Correct
Yes. You need to know what the human performance level is to estimate avoidable bias.

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"?

- 0.75% (average of all four numbers above)
- 0.3% (accuracy of expert #1)
- 0.4% (average of 0.3 and 0.5)
- 0.0% (because it is impossible to do better than this)

Expand

Correct

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

- 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.
- Human-level performance -> Bayes error -> the learning algorithm's 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. You find that a team of ornithologists debating and discussing an image gets an even better 0.1% performance, so you define that as "human-level performance." After working further on your algorithm, you end up with the following: 1 / 1 point

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

Based on the evidence you have, which two of the following four options seem the most promising to try? (Check two options.)

- [ ] Try increasing regularization.

- [x] Try decreasing regularization.

  ✓ Correct

- [ ] Get a bigger training set to reduce variance.
- [x] Train a bigger model to try to do better on the training set.

  ✓ Correct

Expand

✓ Correct
Great, you got all the right answers.

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) **1 / 1 point**

- [x] 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.
- [ ] Get a bigger test set to increase its accuracy.
- [x] Try increasing regularization to reduce overfitting to the dev set.

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

Expand

✓ Correct
Great, you got all the right 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.)

- [ ] This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- [x] If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

  ✓ Correct

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

  ✓ Correct

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? **1 / 1 point**

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

Expand

✓ Correct
Yes. The target has shifted so an updated metric is required.

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? **0 / 1 point**

- (●) Put the new species' images in training data to learn their features.
- ( ) Add pooling layers to downsample features to accommodate the new species.
- ( ) Augment your data to increase the images of the new bird.
- ( ) Split them between dev and test and re-tune.

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. 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.) **0 / 1 point**

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

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

- [ ] With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.
- [ ] 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.
- [x] 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.

Expand

Incorrect
You didn't select all the correct answers