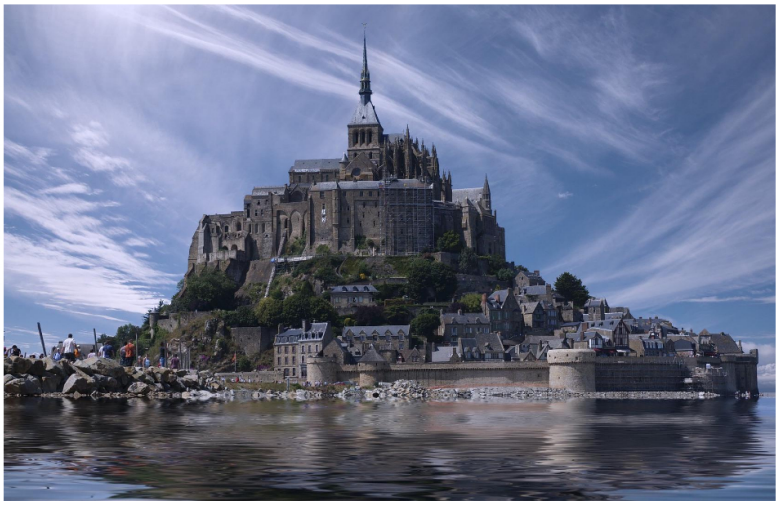
Congratulations! You passed!

Grade received 80% | Latest Submission Grade 53.33% | To pass 80% or higher

Go to next item

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 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 are delighted because this list of criteria will speed development and provide guidance on how to evaluate two different algorithms. True/False?

- (•) False
- ( ) True:

Expand

Correct
Yes. More than one metric expands the choices and tradeoffs you have to decide for each with unknown effects on the other two.

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

suggest they identify the criteria?

- (●) 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.
- ( ) Suggest to them that they focus on whichever criterion is important and then eliminate the other two.

Expand

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

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

- ( ) It isn't. All metrics must be met for the model to be acceptable.
- (●) 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.
- ( ) Knowing the metrics provides input for efficient project planning.

Expand

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

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
6,000,000	1,000,000	3,000,000

(●)

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

( )

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

( )

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

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. You should add the citizens' data to the training set. True/False? **1 / 1 point**

- (●) True
- ( ) False

Expand

Correct
Yes. This will cause the training and dev/test set distributions to become different, however as long as dev/test distributions are the same you are aiming at the same target.

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**

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

  ✓ Correct

- [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

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

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? **0 / 1 point**

- ( ) No, because you do not know what the human performance level is.
- ( ) 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.
- (•) No, because this shows your variance is higher than your bias.

Expand

Incorrect
No. Test accuracy is not given so we can't speak about variance.

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? **0 / 1 point**

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

Expand

Incorrect
You answered when the time elapsed.

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. 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: **0 / 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.)

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

! **This should not be selected**

Expand

Incorrect
You didn't select all the correct answers

11. After running your model with the test set you find it is a 7.0% error compared to a 2.1% error for the dev set and 2.0% for the training set. What can you conclude? (Choose all that apply) **1 / 1 point**

- [x] You have overfitted to 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.
- [ ] You have underfitted to the dev set.
- [x] You should try to get a bigger 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: **0 / 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.
- [x] It is now harder to measure avoidable bias, thus progress will be slower going forward.

✓ Correct

- [ ] 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 $$\leq 0.05%$$

✓ Correct

Expand

✗ Incorrect
You answered when the time elapsed.

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.
- ○ Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ○ Apply regularization to minimize the false negative rate.
- ◉ Ask your team to take into account both accuracy and false negative rate during 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? **0 / 1 point**

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

Expand

✗ Incorrect
You answered when the time elapsed.

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.) **0 / 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.
- [x] 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

trained on less data.

Correct

- [ ] Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.
- [x] Needing two weeks to train will limit the speed at which you can iterate.

Correct

Expand

Incorrect
You didn't select all the correct answers