Congratulations! You passed!

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

Go to next item

1. Problem Statement 0 / 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?

- (●) True:
- ( ) False

Expand

Incorrect
No. 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 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.
- (selected) Suggest to them that they define which criterion is most important. 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**

- (selected) False
- True

Expand

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

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	3,000,000	1,000,000

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

(selected)

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? **0 / 1 point**

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

Expand

Incorrect
No. This would add noise because the images are not from the same cameras which will be used in production.

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**

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

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.
- ☑ 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. 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 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.
- ○ No, because this shows your variance is higher than your bias.

Expand

Correct

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)
- 0.0% (because it is impossible to do better than this)
- 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. 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] Train a bigger model to try to do better on the training set.

Correct

- [ ] Get a bigger training set to reduce variance.
- [x] Try decreasing regularization.

✓ Correct

Expand

✓ Correct
Great, you got all the right 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**

- [ ] Try decreasing regularization for better generalization with the dev set.
- [ ] You have underfitted to the dev set.
- [x] You have overfitted to the dev set.

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

- [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: Human-level performance, 0.10%, Training set error, 0.05%, Dev set error, 0.05%. Which of the following are likely? (Check all that apply.) **1 / 1 point**

- [ ] There is still avoidable bias.
- [x] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.

✓ Correct
Yes. Exceeding human performance means you are close to Bayes error.

- [x] The model has recognized emergent features that humans cannot. (Chess and Go for example)

✓ Correct
Yes. When Google beat the world Go champion, it was recognized that it was making deeper moves than humans.

- [ ] This result is not possible 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. You and your competitor both deliver systems with about the same running time **1 / 1 point**

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?

- (•) 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. 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.
- ( ) Add hidden layers to further refine feature development.
- ( ) Add the new images and split them among train/dev/test.
- ( ) Augment your data to increase the images of the new bird.

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.)

1 / 1 point

- [ ] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.
- [x] Lowering the number of images will reduce training time and likely allow for an acceptable tradeoff between iteration speed and accuracy.

  ✓ **Correct**
  Yes. There is a sweet spot that allows development at a reasonable rate without significant accuracy loss.

- [x] This significantly impacts iteration speed.

  ✓ **Correct**
  Yes. This training time is an absolute constraint on iteration.

Expand

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