Congratulations! You passed!

Grade received 93.33% Latest Submission Grade 93.33% 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

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: **1 / 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
97%	1 sec	3MB

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

Expand

Correct
Correct! This model has the highest test accuracy, the prominent criteria you are looking for, compared with other models, and also has a runtime <10 seconds and memory size < 10MB.

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

- Identifying the optimizing metric informs the team which models they should try first.
- Knowing the metrics provides input for efficient project planning.
- Identifying the metric types sets thresholds for satisficing metrics. This provides explicit evaluation criteria.
- It isn't. All metrics must be met for the model to be acceptable.

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
9,500,000	250,000	250,000

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

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

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**

- False
- True

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 proportionately to the train/dev/test sets. You object because: **1 / 1 point**

- 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.
- The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- The additional data would significantly slow down training time.

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?

- Yes, because this shows your bias is higher than your variance.

- (●) No, because there is insufficient information to tell.
- ( ) 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 want to define what human-level performance is to the city council. Which of the following is the best answer? **1 / 1 point**

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

Expand

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

9. Which of the following statements do you agree with? **1 / 1 point**

- ( ) A learning algorithm's performance can never be better than human-level performance but it can be better than Bayes error.
- ( ) A learning algorithm's performance can never be better than human-level performance nor better than Bayes error.
- (●) A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error.
- ( ) A learning algorithm's performance can be better than human-level performance and better than Bayes error.

Expand

✓ Correct

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.)

- [x] Try decreasing regularization.

✓ Correct

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

✓ Correct

☐ Get a bigger training set to reduce variance.

☐ Try increasing regularization.

Expand

✓ Correct
Great, you got all the right answers.

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.)

☑ You have overfit to the dev set.

✓ Correct

☑ You should try to get a bigger dev set.

✓ Correct

☐ You should get a bigger test set.

☐ You have underfitted to the dev set.

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.)

☑ 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? **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.
- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ( ) Ask your team to take into account both accuracy and false negative rate during 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**

- ( ) Add the new images and split them among train/dev/test.
- (•) Put them into the dev set to evaluate the bias and re-tune.
- ( ) Add hidden layers to further refine feature development.
- ( ) 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**

- [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.

- [x] 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.

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