

✓ Congratulations! You passed!

Grade
received 93.33%

Latest Submission
Grade 93.33%

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 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.

Note: Having three evaluation metrics makes it harder for you to quickly choose between two different algorithms, and will slow down the speed with which your team can iterate. True/False?

- ☒ True
- ☐ False

Expand

✓ Correct

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 define which criterion is most important. Then, set thresholds for the other two.
- ☐ Suggest to them that they focus on whichever criterion is important and then eliminate 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. The essential difference between an optimizing metric and satisficing metrics is the priority assigned by the stakeholders. True/False?

1 / 1 point

- ☐ True

☒ False

✓ 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
- ☒

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

- ☐ Split it among train/dev/test equally.
- ☒ Add it to the training set.
- ☐ 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 proportionately to the train/dev/test sets. You object because:

1 / 1 point

- ☐ The 1,000,000 citizens' data images do not have a consistent $x \rightarrow y$ mapping as the rest of the data.
- ☐ 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 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?

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

- ☐ 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.

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

 Expand

 Correct

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

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. After working on your algorithm you have to decide the next steps. Currently, human-level performance is 0.1%, training is at 2.0% and the dev set is at 2.1%. Which, two of the following four, statements best describe your thought process?

1 / 1 point

☒ Address bias first through a larger model to get closest to human level error.

 Correct

Yes. Selecting the largest difference from (train set error - human level error) and (dev set error - train set error) and reducing bias or variance accordingly is the most productive step.

☐ Get a bigger training set to reduce variance.

☐ Decrease variance via regularization so training and dev sets have similar performance.

☒ Decrease regularization to boost smaller signals.

 Correct

Yes. Bias is higher than variance.

 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

☒ You should try to get a bigger 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 overfitted to the dev set.

✓ **Correct**

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

☐ 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.)

- ☐ With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- ☒ 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**

Loading [MathJax]/jax/output/CommonHTML/jax.js

✓ **Expand**

✓ **Correct**

Great, you got all the right answers.

13. Your system is now very accurate but has a higher false negative rate than the City Council of Peacetopia would like. What is your best next step?

1 / 1 point

- ☐ Expand your model size to account for more corner cases.
- ☒ Reset your "target" (metric) for the team and tune to it.
- ☐ Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ☐ Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.

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

✓ **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

- ☒ 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.

☒ Given a significant budget for cloud GPUs, you could mitigate the training time.

✓ **Correct**

Yes. More resources will allow you to iterate faster.

☒ 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.

✓ **Expand**

✓ **Correct**

Great, you got all the right answers.