Congratulations! You passed!

Next Item

1. **Problem Statement**

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 Peaceropia. The people of Peaceropia have a common characteristic: they are afraid of birds. To ease them, you have **to build an algorithm that will detect any bird flying over Peaceropia** and alert the population.

The City Council gives you a dataset of 10,000,000 images of the skies Peaceropia, taken from the city's security cameras. They are labelled:

- $y = 0$: 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 Peaceropia.

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

Correct

2. After further discussions, the city narrows down its criteria to:

- "We need an algorithm that can let us know a bird is flying over Peaceropia as accurately as possible."
- "We want the trained model to take no more than 10sec to classify a new image."
- "We want the model to fit in 10GB of memory."

If you had the three following models, which one would you choose?

	Test Accuracy	Runtime	Memory size
	97%	1 sec	3MB
	99%	13 sec	9MB
	97%	3 sec	2MB
✓	99%	9 sec	9MB

Correct
Correct! As soon as the runtime is less than 10 seconds you're good. So, you may simply maximize the test accuracy after you made sure the runtime is <10sec.

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

- Accuracy is an optimizing metric; running time and memory size are a satisficing metrics.

Correct

- Accuracy is a satisficing metric; running time and memory size are an optimizing metric.
- 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 satisficing metrics because you have to do sufficiently well on all three for your system to be acceptable.

4. **Structuring your data**

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
✓	9,500,000	250,000	250,000

Correct
Yes.

	Train	Dev	Test
	6,000,000	3,000,000	1,000,000
	3,333,334	3,333,333	3,333,333

5. After setting up your train/dev/test sets, the City Council comes across another 1,000,000 images, called the "citizens' data". Apparently the citizens of Peaceropia are so scared of birds that they volunteered to take pictures of the sky and label them, thus contributing to your dataset. Unfortunately, this data is of slightly different type than the data of all images the City Council had originally given you, but you think it could help your algorithm.

You should not add the citizens' data to the training set, because this will cause the training and dev/test set distributions to become different, thus hurting dev and test set performance. True/False?

- True
- False

Correct
Adding this data to the training set will change the training set distribution, so performance on the dev/test sets will not reflect performance on the distribution of the city's security cameras, which is now different. Dev and test set distributions should come from the city's security cameras, so that the dev set tells you how well the system will perform on the distribution you care about.

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 train set. You disagree because

- 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

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

Correct

- The 1,000,000 citizens' data images do not have a consistent x-y mapping as the rest of the data (similar to the New York City/Detroit housing prices example from lecture).

Un-selected is correct

- A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the dev set.

Un-selected is correct

7. You train a system, and its errors are as follows (error = 100% Accuracy):

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 0.5% training error. Do you agree?

- Yes, because having 0.5% training error shows you have high bias.
- Yes, because this shows your bias is higher than your variance.
- No, because this shows your variance is higher than your bias.
- No, because there is insufficient information to tell.

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:

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" for a proxy (or estimate) for Bayes error, how would you define "human-level performance"?

- 0.3% (because it is impossible to do better than this)
- 0.3% (accuracy of expert #1)

Correct

- 0.4% (average of 0.3 and 0.5)
- 0.75% (average of all four numbers above)

9. Which of the following statements do you agree with?

- A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error.

Correct

- 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 and 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:

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.

Un-selected is correct

- Try decreasing regularization.

Correct

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

Correct

- Try increasing regularization.

Un-selected is correct

11. You also evaluate your model on the test set, and find the following:

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 get a bigger test set.

Un-selected is correct

- You should try to get a bigger dev set.

Correct

- You have underfit to the dev set.

Un-selected is correct

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%

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.

Un-selected is correct

- If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

Correct

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

Correct

- With only 0.09% further progress to make, you should quickly be able to close the remaining gap to 0%

Un-selected is correct

13. It turns out Peaceropia has hired one of your competitors to build a system as well. Your system and theirs both work well, and both of you have roughly human-level performance on the test set. However, your system has higher accuracy! However, when Peaceropia tries out your and your competitor's systems, they conclude they actually like your competitor's system better because it makes fewer false negatives (birds in the sky that you have even false negatives failing to raise an alarm when a bird is in the sky). What should you do?

- Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- Ask your team to take into account both accuracy and false negative rate during development.
- Rethink the appropriate metric for this task, and ask your team to tune to the new metric.

Correct

- Pick false negative rate as the new metric, and use this new metric to drive all further developments.

14. You've hardly beaten your competitor, and your system is now deployed in Peaceropia and is processing the citizens from birds flux 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 having tested on a new type of data.

You have only 1,000 images of the new species of bird. The city expects a better system from you within the next 3 months. Which of these should you do first?

- Use the data you have to define a new evaluation metric (using a new dev/test) that taking into account the new species, and use that to drive further progress for your team.

Correct

- Put the 1,000 images into the training set so as to try to do better on these birds.
- Try data augmentation/data synthesis to get more images of the new type of bird.
- Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.

15. The City Council thinks that having more CAs in the city would help scare off birds. They plan to double the size of the City Council, and they ask you to double your team's size. Your detector. (Bird CA) detectors are just incredibly useful and very profitable. Because of years of working on CA detectors, you have such a huge dataset of 100,000,000 cat images that you've been training on about four weeks. Which of the statements do you agree with? (Check all that apply.)

- If 100,000,000 examples is enough to build a good enough CA detector, you might be better of training with 200,000,000 examples only if it gives a 1.5% improvement in how quickly you can run experiments, even if each model performs a bit worse because it's trained on less data.

Correct

- Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.

Correct

- Training 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.

Un-selected is correct

- Spending two weeks to train will limit the speed at which you can iterate.

Correct