Your grade: 93.33%

Your latest: **93.33%** • Your highest: **93.33%** • To pass you need at least 80%. We keep your highest score.

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?

- ○ False
- ◉ True:

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

2. The city revises 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."

Given models with different accuracies, runtimes, and memory sizes, how would you choose one?

- ○ Accuracy is an optimizing metric, therefore the most accurate model is the best choice.
- ○ Create one metric by combining the three metrics and choose the best performing model.
- ◉ Find the subset of models that meet the runtime and memory criteria. Then, choose the highest accuracy.
- ○ Take the model with the smallest runtime because that will provide the most overhead to increase accuracy.

Correct
Yes. Once you meet the runtime and memory thresholds, accuracy should be maximized.

3. Based on the city's requests, which of the following would you say is true? 1 / 1 point

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

Correct

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
3,333,334	3,333,333	3,333,333

- ( )

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

- ( )

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

Correct
Yes.

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 Peacetopia are so scared of birds that they volunteered to take pictures of the sky and label them, thus contributing these additional 1,000,000 images. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. **1 / 1 point**

Notice that adding this additional data to the training set will make the distribution of the training set different from the distributions of the dev and test sets.

Is the following statement true or false?

"You should not add the citizens' data to the training set, because if the training distribution is different from the dev and test sets, then this will not allow the model to perform well on the test set."

- ( ) True
- (●) False

Correct
False is correct: Sometimes we'll need to train the model on the data that is available, and its distribution may not be the same as the data that will occur in production. Also, adding training data that differs from the dev set may still help the model improve performance on the dev set. What matters is that the dev and test set have the same distribution.

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 dev set. You object because: (Choose all that apply) **1 / 1 point**

- [ ] The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- [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
Yes. Adding a different distribution to the dev set will skew bias.

- [ ] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.
- [x] The dev set no longer reflects the distribution of data (security cameras) you most care about.

Correct
Yes. The performance of the model should be evaluated on the same distribution of images it will see in production.

7. Human performance for identifying birds is < 1%, training set error is 5.2% and dev set error is 7.3%. Which of the options below is the best next step? **1 / 1 point**

- ( ) Validate the human data set with a sample of your data to ensure the images are of sufficient quality.
- ( ) Try an ensemble model to reduce bias and variance.
- ( ) Get more data or apply regularization to reduce variance.
- (●) Train a bigger network to drive down the >4.0% training error.

Correct
Yes. Avoidable bias is >4.2% which is larger than the 2.1% variance.

8. 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 / 1 point**

- ( ) The best performance of a specialist (ornithologist) or possibly a group of specialists.
- (●) The performance of their volunteer amateur ornithologists.
- ( ) The performance of the average citizen of Peacetopia.
- ( ) The performance of the head of the City Council.

Incorrect
No. These are undoubtedly better than a random citizen but not as good as a specialist.

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 and 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.

✓ **Correct**

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**

- ☑ Decrease regularization to boost smaller signals.

✓ **Correct**
Yes. Bias is higher than variance.

- ☑ 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.

- ☐ Decrease variance via regularization so training and dev sets have similar performance.
- ☐ Get a bigger training set to reduce variance.

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 should try to get a bigger dev set.

✓ **Correct**

- ☐ You have underfitted to the dev set.
- ☑ You have overfit to the dev set.

✓ **Correct**

- ☐ You should get a bigger test set.

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**

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

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

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.
- ○ Ask your team to take into account both accuracy and false negative rate during development.
- ○ 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.

✓ **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 data is being tested on a new type of data. **1 / 1 point**

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 set) taking into account the new species, and use that to drive further progress for your team.
- ( ) Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.
- ( ) Try data augmentation/data synthesis to get more images of the new type of bird.
- ( ) Put the 1,000 images into the training set so as to try to do better on these birds.

Correct

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.) 1 / 1 point

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

Correct

- [ ] 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.

Correct

- [x] Needing two weeks to train will limit the speed at which you can iterate.

Correct