Your grade: 80%

Your latest: **80%** • Your highest: **80%** • 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. 0 / 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:

✗ **Incorrect**
No. More than one metric expands the choices and tradeoffs you have to decide for each with unknown effects on the other two.

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.
- ( ) Take the model with the smallest runtime because that will provide the most overhead to increase accuracy.
- (●) Find the subset of models that meet the runtime and memory criteria. Then, choose the highest accuracy.
- ( ) Create one metric by combining the three metrics and choose the best performing model.

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, 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.
- ◉ Accuracy is an optimizing metric; running time and memory size are satisfying metrics.
- ○ Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.
- ○ Accuracy is a satisfying metric; running time and memory size are an optimizing metric.

Correct

4. You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal? 1 / 1 point

- ○ The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.
- ◉ With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.
- ○ The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.
- ○ The emphasis on the training set will allow us to iterate faster.

Correct
Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

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

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 to the dev set. You object because: (Choose all that apply) 0 / 1 point

- ☑ The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.

This should not be selected
No. The important issue is mixing distributions.

- ☑ 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.
- ☑ 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. 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 having a 4.0% training error shows you have a high bias.

- (●) No, because there is insufficient information to tell.
- ( ) No, because this shows your variance is higher than your bias.
- ( ) Yes, because this shows your bias is higher than your variance.

Correct

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"? **1 / 1 point**

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

Correct
Yes. This is the peak of human performance in this task.

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

- ( ) A learning algorithm's performance can be better than human-level performance and better than Bayes error.
- ( ) 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.

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.)

- [ ] Get a bigger training set to reduce variance.
- [ ] Try increasing regularization.
- [x] Train a bigger model to try to do better on the training set.

Correct

- [x] Try decreasing regularization.

Correct

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 get a bigger test set.
- [x] You should try to get a bigger dev set.

Correct

- [ ] You have underfitted to the dev set.
- [x] You have overfit to the dev set.

Correct

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.)

- [x] 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.
- [x] 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%

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? 0 / 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.

Incorrect
No. This choice also points to the incorrect target.

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.
- ( ) 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.

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. 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.

- [ ] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.
- [x] This significantly impacts iteration speed.

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

- [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.