Your grade: 91.66%

Your latest: **91.66%** • Your highest: **91.66%** • To pass you need at least 80%. We keep your highest score.

Next item →

1. This example is adapted from a real production application, with details modified for confidentiality. 1 / 1 point

You are a renowned researcher in the City of Peacetopia. The residents of Peacetopia share a unique characteristic: they are afraid of birds. To protect them, you are tasked with developing an algorithm that will detect any bird flying over Peacetopia and alert the population.

The City Council provides you with a dataset of 10,000,000 images of the sky above Peacetopia, captured by

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 create an algorithm capable of classifying new images taken by security cameras in Peacetopia. You have several decisions to make regarding the evaluation metric and how to structure your data into train/dev/test sets.

The City Council specifies that they want an algorithm that:

1. Has high accuracy.
2. Operates quickly and takes only a short time to classify a new image.
3. Requires minimal memory, allowing it to run on a small processor attached to various security cameras.

True or False: You discuss with them the need for a singular evaluation metric to guide development.

- (●) True
- ( ) False

Correct
Choosing one primary metric simplifies decision-making and enhances development speed, allowing for clearer comparisons between different models.

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 seconds to classify a new image."
- "We want the model to fit in 10MB of memory."
- "We require a minimum of 98% test accuracy."

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

- (●)

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

- ( )

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

- ( )

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

Correct
The runtime is less than 10 seconds, and the accuracy meets the minimum 98% requirement.

3. Based on the context of a city's data analysis project, **which of the following statements is true regarding the metrics used?** 1 / 1 point

- (●) Accuracy is an optimizing metric; running time and memory size are satisficing metrics.
- ( ) Accuracy is a satisficing metric; running time and memory size are an optimizing metric.
- ( ) 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.
- ( ) Accuracy, running time, and memory size are all optimizing metrics because you want to do well on all three.

Correct
Accuracy is a metric we aim to maximize, while running time and memory size have acceptable thresholds.

4. With 10,000,000 data points, what is the best option for train/dev/test splits? 1 / 1 point

- ( ) train - 60%, dev - 30%, test - 10%
- (●) train - 95%, dev - 2.5%, test - 2.5%
- ( ) train - 33.3%, dev - 33.3%, test - 33.3%
- ( ) train - 60%, dev - 10%, test - 30%

Correct
The size of the data set allows for effective bias and variance evaluation with smaller data sets.

5. After setting up your train/dev/test sets, the City Council comes across another 1,000,000 images, called the 1 / 1 point

"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 have a different distribution from the images the City Council originally provided, but you think they could help your algorithm. Notice that adding this additional data to the training set will make the distribution of the training set different from the distribution of the dev and test sets.

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
Sometimes we'll need to train the model on available data, even if its distribution differs from the data that will occur in production. Adding training data that differs from the dev set may still help the model improve performance. What matters is that the dev and test sets have the same distribution.

6. One member of the City Council wants to add 1,000,000 citizen data images to the test set. Your original data is from security cameras, and you object because: (Choose all that apply) **0.8 / 1 point**

- ☑ The 1,000,000 citizen data images do not have a consistent input-output relationship as the security camera data.

This should not be selected
The important issue is mixing distributions, rather than minor differences in the input-output relationship.

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

Correct
Yes. The test set must reflect real world 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. Mixing distributions between dev and test sets will cause issues with measuring model performance.

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

7. Human performance for identifying birds is < 1%, training set error is 5.2%, and dev set error is 7.3%. **1 / 1 point**

Which of the options below is the best next step?

- ( ) Get more data or apply regularization to reduce variance.
- ( ) Try an ensemble model to reduce bias and variance.
- ( ) Validate the human data set with a sample of your data to ensure the images are of sufficient quality.
- (•) Train a bigger network to reduce the 5.2% training error.

Correct
Avoidable bias is 4.2%, which is larger than the 2.1% variance, so reducing bias is the priority.

8. If your goal is to use "human-level performance" as an estimate for Bayes error in a bird species identification task, how would you define "human-level performance"? **1 / 1 point**

- (•) The best performance of a specialist (ornithologist) or a group of specialists.
- ( ) The performance of the head of the City Council.

- ○ The performance of the average citizen.
- ○ The performance of volunteer amateur ornithologists.

Correct
This represents the peak of human performance in this task, closely approximating Bayes error.

9. Which of the below shows the optimal order of accuracy from worst to best? 1 / 1 point

- ○ Human-level performance -> Bayes error -> the learning algorithm's performance.
- ○ The learning algorithm's performance -> Bayes error -> human-level performance.
- ○ The learning algorithm's performance -> human-level performance -> Bayes error.
- ◉ Human-level performance -> the learning algorithm's performance -> Bayes error.

Correct
A learning algorithm can outperform human-level performance, but it cannot surpass the Bayes error, which is the theoretical limit of accuracy.

10. Which of the following best describes the **most effective next step in your project**, given the following performance metrics? 1 / 1 point

- Human-level performance: 0.1%
- Training set error: 2.0%
- Dev set error: 2.1%

- ○ Deploy the model to target devices to evaluate against satisficing metrics.
- ◉ Prioritize actions to decrease bias by increasing model complexity, as the training error significantly

exceeds human-level performance.

- ○ Evaluate the test set to determine the variance.
- ○ Continue tuning until the training set error matches human-level performance, focusing solely on the optimizing metric.

Correct
Yes, addressing the largest performance gap (between human-level and training error) is the most efficient strategy.

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
A larger dev set can provide a more reliable performance estimate and help address overfitting issues.

- ☐ You have underfitted to the dev set.
- ☐ You should get a bigger test set.
- ☑ You have overfit to the dev set.

Correct
The much higher test set error compared to the dev set indicates overfitting to the dev set.

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%. Which of the following are likely? (Check all that apply.) 1 / 1 point

- [ ] There is still avoidable bias.
- [x] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.

Correct
Exceeding human performance means you are close to Bayes error, making bias identification difficult.

- [x] The model has recognized complex, emergent features that humans may not readily perceive. (Chess and Go, for example).

Correct
In domains like Chess and Go, AI models have demonstrated the ability to discover and utilize strategies beyond typical human comprehension.

- [ ] This result is not possible since it should not be possible to surpass human-level performance.

13. It turns out Peacetopia has hired one of your competitors to build a system as well. Your system and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! 1 / 1 point

Still, when Peacetopia tries out both your system and your competitor's system, they conclude they actually 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?

- (●) Rethink the appropriate metric for this task, and ask your team to tune to the new metric.
- ( ) Look at all the models you've developed during the development process and find the one with the lowest false negative error 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.

Correct
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 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?

- ( ) Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split
- ( ) Use the data you have to define a new evaluation metric (using a new dev/test set) that accounts for the new species, and use that metric to guide further improvements.
- (●) 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.

Incorrect
Data augmentation can be a useful technique, but it should follow the adjustment of the evaluation metric to ensure you're measuring progress accurately.

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. 1 / 1 point

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.)

- [ ] 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 ~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
A smaller dataset can expedite training and allow for more iterations, potentially leading to a satisfactory model faster.

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

Correct
The long training time constrains how quickly you can test and refine models.

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

Correct
Enhanced computational resources can reduce training time and improve productivity.