Your grade: 100%

Your latest: **100%** • Your highest: **100%**
To pass you need at least 80%. We keep your highest score.

Next item →

1. This example is inspired by a real-world application, with details modified for privacy. **1 / 1 point**

In the City of Peacetopia, you are a celebrated researcher tasked with a critical mission. The entire population shares a phobia of birds, and your job is to create an algorithm that identifies any birds flying over the city, alerting residents promptly.

You are given 10,000,000 images of the sky from the city's security cameras, labeled as:

- y = 0: There is no bird on the image
- y = 1: There is a bird on the image

Your mission is to develop an algorithm to classify new images from security cameras accurately. This involves crucial decisions on the

evaluation metric and how to organize your data into train/dev/test sets.

The City Council desires an algorithm that:

1. Achieves high accuracy.
2. Quickly classifies new images.
3. Uses minimal memory to be compatible with small processors in various security cameras.

True or False: You acknowledge that having multiple evaluation metrics may complicate the decision-making process and slow down iteration speed.

- ○ False
- ◉ True

Correct
While it's important to consider various performance aspects, focusing on a single evaluation metric simplifies decisions and accelerates the development cycle, enabling faster iterations and optimizations.

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
99%	13 sec	9MB

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

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. **True or False:** The essential difference between an optimizing metric and satisficing metrics is the priority assigned by the stakeholders. **1 / 1 point**

False

- ◯ True

Correct
Satisficing metrics have specific thresholds that need to be met, while an optimizing metric is one we aim to maximize beyond any set threshold.

4. You propose a 95% / 2.5% / 2.5% for train / dev / test splits to the City Council. They ask for your reasoning. **1 / 1 point**

Which of the following **best justifies your proposal**, given that the total data set contains 10,000,000 data points?

- ◯ The emphasis on the training set provides the most accurate model, supporting the memory and processing efficiency.
- ◉ With a dataset comprising 10,000,000 individual samples, 2.5% represents 250,000 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
The purpose of dev and test sets is fulfilled even with smaller percentages of the data in large datasets, allowing for effective evaluation of model performance.

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 have a different distribution from the images the City Council **1 / 1 point**

originally provided, but you think they could help your algorithm. Should you add this data to the training set?

- (●) Yes.
- ( ) No.

Correct
This will cause the training and dev/test set distributions to become different. However, as long as the dev/test distributions are the same, you are aiming at the same target.

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) **1 / 1 point**

- [x] 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.

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

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

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 to reduce the 4.0% training error. Do you agree?

- ○ Yes, because having a 4.0% training error shows you have a high bias.
- ○ Yes, because this shows your bias is higher than your variance.
- ◉ No, because there is insufficient information to tell.
- ○ No, because this shows your variance is higher than your bias.

Correct
Without knowing the human performance level, we can't determine the bias accurately.

8. You want to define "human-level performance" for a bird species identification project to present to the city council. Which of the following is the best way to define it? **1 / 1 point**

- ○ The average performance of all the city's ornithologists (0.5%).
- ○ The average performance of regular citizens of Peacetopia (1.2%).
- ○ The average of all recorded error rates (0.66%, ornithologists and citizens).
- ◉ The performance of the city's best ornithologist (0.3% error rate).

Correct
The best human performance, represented by the lowest error rate, is the closest practical estimate of Bayes' error.

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

Correct
This statement accurately reflects the relationship between the learning algorithm, human-level performance, and Bayes error.

10. After evaluating your algorithm's performance, you need to determine the next steps. **1 / 1 point**

The metrics are:

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

Which **two** of the following statements best describe **your most effective thought process**?

- ☐ Increase the training dataset size to reduce variance.

- [x] Address bias by increasing model complexity to reduce the gap between training error and human-level performance.

Correct
Yes, the largest difference (training error - human-level error) indicates high bias, necessitating a focus on increasing model complexity.

- [x] Decrease regularization to reduce bias and allow the model to capture finer patterns.

Correct
Yes, given the high bias, reducing regularization can help the model fit the training data better.

- [ ] Decrease variance through increased regularization to align training and dev set performance.

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.)

- [x] 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.

- [ ] You have underfitted to the dev set.
- [x] 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 should get a bigger test 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**

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

- [ ] There is still avoidable bias.
- [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.

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?

- ○ Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ◉ 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.
- ○ 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. 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. There are only 1,000 images of the new species. The city expects a better system from you within the next 3 months.

1 / 1 point

Which of these should you do first?

- ◉ Augment your data to increase the number of images of the new bird species.

- ○ Add pooling layers to downsample features to accommodate the new species.
- ○ Put the new species' images in training data to learn their features.
- ○ Split them between dev and test and re-tune.

Correct
Generating more training examples through data augmentation is the most effective first step to address the limited data for the new species.

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.

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.
- ☑ 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] 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.

- [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.