Final Quiz

Congratulations! You passed!

Grade received	Latest Submission Grade	To pass
86.66%	86.67%	80% or higher

Go to next item

1. Problem Statement 0 / 1 point

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 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 are delighted because this list of criteria will speed development and provide guidance on how to evaluate two different algorithms. True/False?

- (●) True:
- ( ) False

Expand

Incorrect
No. The goal is to have one metric that focuses the development effort and increases iteration velocity.

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 sec to classify a new image."
- "We *want* the model to fit in 10MB of memory."

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
97%	3 sec	2MB

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

Expand

Correct

Correct! This model has the highest test accuracy, the prominent criteria you are looking for, compared with other models, and also has a runtime <10 seconds and memory size < 10MB.

3. Which of the following best answers why it is important to identify optimizing and satisficing metrics? 1 / 1 point

- Identifying the optimizing metric informs the team which models they should try first.
- Identifying the metric types sets thresholds for satisficing metrics. This provides explicit evaluation criteria.
- It isn't. All metrics must be met for the model to be acceptable.
- Knowing the metrics provides input for efficient project planning.

Expand

Correct
Yes. Thresholds are essential for evaluation of key use case constraints.

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

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

Expand

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

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. Which of the following is the best use of that additional data?

0 / 1 point

- ○ Add it to the training set.
- ○ Add it to the dev set to evaluate how well the model generalizes across a broader set.
- ◉ Split it among train/dev/test equally.
- ○ Do not use the data. It will change the distribution of any set it is added to.

Expand

✗ **Incorrect**
No. This would add noise because the images are not from the same cameras which will be used in production.

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

- ☑ 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.

- ☑ 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.

- ☐ A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.
- ☐ The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.

Expand

✓ **Correct**
Great, you got all the right answers.

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?

- ○ No, because this shows your variance is higher than your bias.
- ◉ No, because there is insufficient information to tell.
- ○ 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.

Expand

✓ 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 performance of the head of the City Council.
- ◉ 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.

Expand

✓ 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 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 nor 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.

Expand

Correct

10. Which of the following best expresses how to evaluate the next steps in your project when your results for human-level performance, train, and dev set error are 0.1%, 2.0%, and 2.1% respectively? 1 / 1 point

- ◉ Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.
- ○ Port the code to the target devices to evaluate if your model meets or exceeds the satisficing metrics.
- ○ Evaluate the test set to determine the magnitude of the variance.
- ○ Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.

Expand

Correct
Yes. Always choose the area with the biggest opportunity for improvement.

11. After running your model with the test set you find it is a 7.0% error compared to a 2.1% error for the dev set and 2.0% for the training set. What can you conclude? (Choose all that apply)

1 / 1 point

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

  ✓ **Correct**
  Yes. The dev set performance versus the test set indicates it is overfitting.

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

  ✓ **Correct**
  Yes. The dev set performance versus the test set indicates it is overfitting.

- [ ] Try decreasing regularization for better generalization with the dev set.

Expand

✓ **Correct**
Great, you got all the right answers.

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

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

  ✓ **Correct**
  Yes. Exceeding human performance means you are close to Bayes error.

- [ ] This result is not possible since it should not be possible to surpass human-level performance.
- [x] The model has recognized emergent features that humans cannot. (Chess and Go for example)

  ✓ **Correct**
  Yes. When Google beat the world Go champion, it was recognized that it was making deeper moves than humans.

- [ ] There is still avoidable bias.

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! However, when Peacetopia tries out your and your competitor's systems, 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?

1 / 1 point

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

Expand

Correct

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?

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

Expand

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**

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

  ✓ **Correct**

- [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**

- [ ] 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.

Expand

Correct
Great, you got all the right answers.