Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Go to next item

1. Problem Statement 1 / 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?

- (●) False
- ( ) True:

Expand

Correct
Yes. 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?

- 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.
- Accuracy is an optimizing metric, therefore the most accurate model is the best choice.

Expand

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, 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.
- Accuracy is an optimizing metric; running time and memory size are satisfying metrics.

Expand

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 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 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 emphasis on the training set will allow us to iterate faster.

Expand

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. Which of the following is the best use of that additional data? **1 / 1 point**

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

Expand

Correct
Yes. It is not a problem to have different training and dev distributions. Different dev and test distributions would be an issue.

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**

- [ ] 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. Adding a different distribution to the dev set will skew bias.

- [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.

- [ ] 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.
- ( ) Yes, because this shows your bias is higher than your variance.
- ( ) Yes, because having a 4.0% training error shows you have a high bias

- [ ] Yes, because having 2% training error shows you have a high bias.
- [x] No, because there is insufficient information to tell.

Expand

Correct

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? **1 / 1 point**

- [ ] The average of regular citizens of Peacetopia (1.2%).
- [ ] The average of all the numbers above (0.66%).
- [ ] The average performance of all their ornithologists (0.5%).
- [x] The performance of their best ornithologist (0.3%).

Expand

Correct
Yes. The best human performance is closest to Bayes' error.

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 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.
- [x] A learning algorithm's performance can be better than human-level performance but it can never 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**

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

Expand

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

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

✓ **Correct**

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

✓ **Correct**

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**

- [ ] 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.

- [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.

- [ ] There is still avoidable bias.

Expand

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

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**

- ( ) Ask your team to take into account both accuracy and false negative rate during development.
- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.

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.

Rethink the appropriate metric for this task, and ask your team to tune to the new metric.

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?

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.

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**

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

Needing two weeks to train will limit the speed at which you can iterate.

Correct

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.

Expand

Correct

Great, you got all the right answers.