Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Go to next item

1. This example is adapted from a real production application, but with details disguised to protect confidentiality. **1 / 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:

Expand

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

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?

- ○ 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.
- ○ 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.

Expand

Correct
Yes. Once you meet the runtime and memory thresholds, accuracy should be maximized.

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

- ○ Knowing the metrics provides input for efficient project planning.
- ○ 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.

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

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. You should add the citizens' data to the training set. True/False? **1 / 1 point**

- ○ False
- ◉ True

Expand

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 test set. You object because: **1 / 1 point**

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

  ✓ Correct

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

Expand

Correct
Great, you got all the right answers.

7. You train a system, and the train/dev set errors are 3.5% and 4.0% respectively. You decide to try regularization to close the train/dev accuracy gap. Do you agree? **0 / 1 point**

- ( ) Yes, because having a 4.0% training error shows you have a high bias.
- ( ) No, because this shows your variance is higher than your bias.
- ( ) No, because you do not know what the human performance level is.
- (•) Yes, because this shows your bias is higher than your variance.

Expand

Incorrect
No. Test accuracy is not given so we can't speak about variance.

8. You ask a few people to label the dataset so as to find out what is human-level performance. You find the following levels of accuracy: **1 / 1 point**

Bird watching expert #1	0.3% error
Bird watching expert #2	0.5% error
Normal person #1 (not a bird watching expert)	1.0% error
Normal person #2 (not a bird watching expert)	1.2% error

If your goal is to have "human-level performance" be a proxy (or estimate) for Bayes error, how would you define "human-level performance"?

- (•) 0.3% (accuracy of expert #1)
- ( ) 0.75% (average of all four numbers above)
- ( ) 0.0% (because it is impossible to do better than this)
- ( ) 0.4% (average of 0.3 and 0.5)

Expand

Correct

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

- (●) Human-level performance -> the learning algorithm's performance -> Bayes error.
- ( ) The learning algorithm's performance -> human-level performance -> Bayes error.
- ( ) Human-level performance -> Bayes error -> the learning algorithm's performance.
- ( ) The learning algorithm's performance -> Bayes error -> human-level performance.

Expand

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

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**

- ( ) 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.
- ( ) Port the code to the target devices to evaluate if your model meets or exceeds the satisficing metrics.
- (●) Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.

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) **0 / 1 point**

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

This should not be selected
No. The dev set performance versus the test set indicates it is overfitting.

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

Expand

Incorrect
You didn't select all the correct 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] 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.

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

performance.

- [ ] 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**
Yes. Exceeding human performance means you are close to Bayes error.

Expand

Correct
Great, you got all the right answers.

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? **1 / 1 point**

- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.
- (•) Reset your "target" (metric) for the team and tune to it.
- ( ) Expand your model size to account for more corner cases.
- ( ) Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.

Expand

Correct
Yes. 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 / 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?

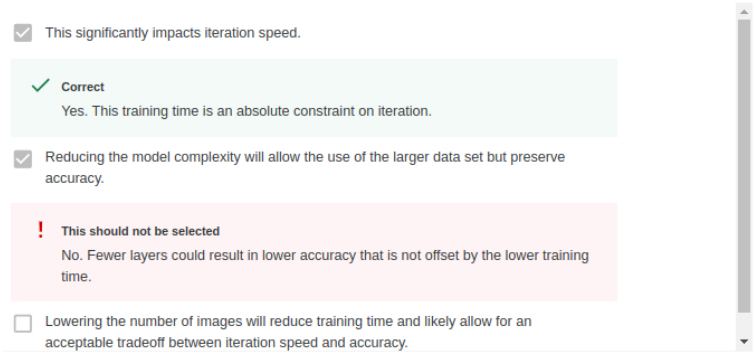
- ( ) Try data augmentation/data synthesis to get more images of the new type of bird.
- (•) Use the data you have to define a new evaluation metric (using a new dev/test set) taking into account the new species, and use that to drive further progress for your team.
- ( ) Add the 1,000 images into your 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.

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. You have a huge dataset of 100,000,000 **0 / 1 point**

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.)



Expand

Incorrect
You didn't select all the correct answers