Congratulations! You passed!

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

Retake the assignment in 23h 34m

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 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.

Note: Having three evaluation metrics makes it harder for you to quickly choose between two different algorithms, and will slow down the speed with which your team can iterate. True/False?

- (●) True
- ( ) False

Expand

Correct

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.
- Accuracy is an optimizing metric, therefore the most accurate model is the best choice.
- Create one metric by combining the three metrics and choose the best performing model.
- 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? 0 / 1 point

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

Expand

Incorrect
No. There may be some benefit to project planning but the most important value is in evaluation.

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

- train - 33.3%, dev - 33.3%, test - 33.3%
- train - 95%, dev - 2.5%, test - 2.5%
- train - 60%, dev - 30%, test - 10%
- 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. Which of the following is the best use of that additional data? **1 / 1 point**

- ○ Add it to the dev set to evaluate how well the model generalizes across a broader set.
- ◉ Add it to the training 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

✓ **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.
- ☑ 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 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- ☑ 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.

Expand

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

7. Human performance for identifying birds is < 1%, training set error is 5.2% and dev set error is 7.3%. Which of the options below is the best next step? **1 / 1 point**

- ○ Get more data or apply regularization to reduce 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 drive down the >4.0% training error.
- ○ Try an ensemble model to reduce bias and variance.

Expand

✓ **Correct**
Yes. Avoidable bias is >4.2% which is larger than the 2.1% variance.

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 average citizen of Peacetopia.
- The performance of their volunteer amateur ornithologists.
- The performance of the head of the City Council.
- (selected) The best performance of a specialist (ornithologist) or possibly a group of specialists.

Expand

Correct
Yes. This is the peak of human performance in this task.

9. A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error. True/False? **1 / 1 point**

- False.
- (selected) True.

Expand

Correct
Yes. By definition, human level error is worse 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**

- (selected) 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.
- Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.
- 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've now also run your model on the test set and find that it is a 7.0% error compared to a 2.1% error for the dev set. What should you do? (Choose all that apply) **1 / 1 point**

- [x] Increase the size of the dev set.

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

- [ ] Get a bigger test set to increase its accuracy.
- [x] Try increasing regularization to reduce overfitting to the 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

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.) **0 / 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.

- [x] There is still avoidable bias.

This should not be selected
No. Exceeding human performance makes the identification of avoidable bias very challenging.

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

Expand

Incorrect
You didn't select all the correct answers

13. It turns out Peacetopia has hired one of your competitors to build a system as well. You and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! Still, when Peacetopia tries out both systems, they conclude they 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.
- (•) Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.
- ( ) Apply regularization to minimize the false negative rate.
- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.

Expand

Correct
Yes. 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. Which of these should you do first? **0 / 1 point**

- ( ) Augment your data to increase the images of the new bird.
- (•) Add pooling layers to downsample features to accommodate the new species.
- ( ) Split them between dev and test and re-tune.
- ( ) Put the new species' images in training data to learn their features.

Expand

Incorrect
No. Pooling layers won't reduce the features in a meaningful way to learn the new species.

No. Pooling layers won't reduce the features in a meaningful way to learn 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. 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.)

1 / 1 point

- [x] Lowering the number of images will reduce training time and likely allow for an acceptable tradeoff between iteration speed and accuracy.

  ✓ **Correct**
  Yes. There is a sweet spot that allows development at a reasonable rate without significant accuracy loss.

- [x] This significantly impacts iteration speed.

  ✓ **Correct**
  Yes. This training time is an absolute constraint on iteration.

- [ ] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.

Expand

Correct
Great, you got all the right answers.