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?

- ○ 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.
- ○ Create one metric by combining the three metrics and choose the best performing model.

Expand

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

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

- ○ True
- ◉ False

Expand

Correct
Yes. Satisficing metrics have thresholds for measurement and an optimizing metric is unbounded.

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

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**

- ○ Split it among train/dev/test equally.
- ◉ Add it to the training set.
- ○ Add it to the dev set to evaluate how well the model generalizes across a broader set.
- ○ 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**

- ☑ 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.

- ☐ 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.
- ☑ 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**

- ◉ Train a bigger network to drive down the >4.0% training error.
- ○ Validate the human data set with a sample of your data to ensure the images are of sufficient quality.
- ○ Get more data or apply regularization to reduce variance.
- ○ 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. 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”?

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

Expand

✓ Correct

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 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.
- ○ A learning algorithm's performance can be better than human-level performance and better than Bayes error.

Expand

✓ Correct

10. After working on your algorithm you have to decide the next steps. Currently, human-level performance is 0.1%, training is at 2.0% and the dev set is at 2.1%. Which statement below best describes your thought process? **1 / 1 point**

- ☐ Decrease variance via regularization so training and dev sets have similar performance.
- ☐ Get a bigger training set to reduce variance.
- ☑ Address bias first through a larger model to get closest to human level error.

  ✓ Correct
  Yes. Selecting the largest difference from (train set error - human level error) and (dev set error - train set error) and reducing bias or variance accordingly is the most productive step.
- ☑ Decrease regularization to boost smaller signals.

  ✓ Correct
  Yes. Bias is higher than variance.

Expand

✓ Correct
Great, you got all the right answers.

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

✓ Correct

- [ ] You should get a bigger test set.
- [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 true? (Check all that apply.) **1 / 1 point**

- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- [x] You are close to Bayes error and possible overfitting.

✓ Correct
Yes. By definition, Bayes error cannot be exceeded except for overfitting.

- [ ] This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- [x] All or almost all of the avoidable bias has been accounted for.

✓ Correct
Yes. Exceeding human performance makes the identification of avoidable bias very challenging.

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. 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**

- ○ Apply regularization to minimize the false negative rate.
- ◉ Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.
- ○ 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.

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? **1 / 1 point**

you do first.

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

Expand

Correct
Yes. A sufficient number of images is necessary to account 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. 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.