Congratulations! You passed!

Grade received 80% **To pass** 80% or higher

Go to next item

Bird Recognition in the City of Peacetopia (Case Study)

Latest Submission Grade 80%

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.
- ( ) 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.
- ( ) 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 is an optimizing metric; running time and memory size are satisfying metrics.
- ( ) Accuracy is a satisfying metric; running time and memory size are an optimizing metric.
- ( ) Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.
- ( ) 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.

Expand

Correct

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

Expand

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

5. After setting up your train/dev/test sets, the City Council comes across another 1,000,000 images, called the "citizens' data". Apparently the citizens of Peacetopia are so scared of birds that they volunteered to take pictures of the sky and label them, thus contributing these additional 1,000,000 images. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. — 1 / 1 point

Notice that adding this additional data to the training set will make the distribution of the training set different from the distributions of the dev and test sets.

Is the following statement true or false?

"You should not add the citizens' data to the training set, because if the training distribution is different from the dev and test sets, then this will not allow the model to perform well on the test set."

- ( ) True
- (•) False

Expand

Correct
False is correct: Sometimes we'll need to train the model on the data that is available, and its distribution may not be the same as the data that will occur in production. Also, adding training data that differs from the dev set may still help the model improve performance on the dev set. What matters is that the dev and test set have the same distribution.

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] 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 (similar to the New York City/Detroit housing prices example from the lecture).
- [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.

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.
- ( ) Try an ensemble model to reduce bias and 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.

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"?

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

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

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. 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? — 0 / 1 point

- [ ] Decrease variance via regularization so training and dev sets have similar performance.
- [x] 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.
- [ ] Get a bigger training set to reduce variance.

Expand

Incorrect
You didn't select all the correct answers

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

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

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

- [ ] You have overfitted to the 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 true? (Check all that apply.) — 1 / 1 point

- [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.

- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%

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

- ( ) Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- ( ) 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.
- (•) 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 model is being tested on a new type of data. Which of these should you do first? — 0 / 1 point

- (•) Add the new images and split them among train/dev/test.
- ( ) Augment your data to increase the images of the new bird.
- ( ) Put them into the dev set to evaluate the bias and re-tune.
- ( ) Add hidden layers to further refine feature development.

Expand

Incorrect
No. The number of new images is too small to make a difference.

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.