Congratulations! You passed!

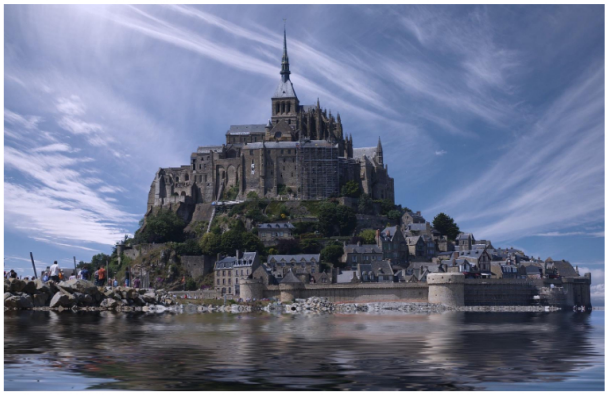
Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Retake the assignment in 20h 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 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?

- ○ True:
- ◉ False

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 asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria? 1 / 1 point

- ○ Suggest to them that they focus on whichever criterion is important and then eliminate the other two.
- ○ Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.
- ◉ Suggest to them that they define which criterion is most important. Then, set thresholds for the other two.

Expand

Correct
Yes. The thresholds provide a way to evaluate models head to head.

3. The essential difference between an optimizing metric and satisficing metrics is the priority assigned by the stakeholders. True/False? 0 / 1 point

- ○ False

True

Expand

Incorrect
No. Stakeholders must define thresholds for satisficing metrics, leaving the optimizing metric unbounded.

4. Structuring your data 1 / 1 point

Before implementing your algorithm, you need to split your data into train/dev/test sets. Which of these do you think is the best choice?

Train	Dev	Test
6,000,000	1,000,000	3,000,000

Train	Dev	Test
3,333,334	3,333,334	3,333,334

Train	Dev	Test
9,500,000	250,000	250,000

Train	Dev	Test
6,000,000	3,000,000	1,000,000

Expand

Correct
Yes.

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? 0 / 1 point

Do not use the data. It will change the distribution of any set it is added to.

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.

Expand

Incorrect
No. This would add noise because the images are not from the same cameras which will be used in production.

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.

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

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 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. 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? 1 / 1 point

No. because you do not know what the human performance level is.

Yes. because this shows your bias is higher than your variance.

- No, because this shows your variance is higher than your bias.
- Yes, because having a 4.0% training error shows you have a high bias.

Expand

Correct
Yes. You need to know what the human performance level is to estimate avoidable bias.

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

Expand

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

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

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

- [x] Decrease regularization to boost smaller signals.

Correct
Yes. Bias is higher than variance.

- [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.

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

Expand

Correct
Great, you got all the right answers.

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.

- [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.
- [ ] Get a bigger test set to increase its accuracy.

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**

- [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.

- [x] You are close to Bayes error and possible overfitting.

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

- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- [ ] This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.

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**

- (•) Reset your "target" (metric) for the team and tune to it.
- ( ) 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.
- ( ) Expand your model size to account for more corner cases.

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 model 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**

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

- [ ] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.
- [x] This significantly impacts iteration speed.

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

- [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.

Expand

Correct

Correct

Great, you got all the right answers.