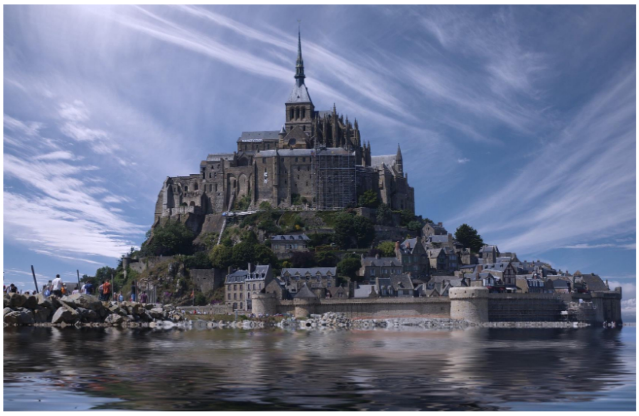
Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

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. After further discussions, the city narrows down 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."

If you had the three following models, which one would you choose?

◉

Test Accuracy	Runtime	Memory size
98%	9 sec	9MB

○

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

○

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

○

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

Expand

Correct
Correct! This model has the highest test accuracy, the prominent criteria you are looking for, compared with other models, and also has a runtime <10 seconds and memory size < 10MB.

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

1 / 1 point

- (•) False
- ( ) True

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

- [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.

Expand

Correct
Great, you got all the right answers.

7. You train a system, and its errors are as follows (error = 100%-Accuracy): 1 / 1 point

Training set error	4.0%
Dev set error	4.5%

This suggests that one good avenue for improving performance is to train a bigger network so as to drive down the 4.0% training error. Do you agree?

- ○ 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.
- ◉ No, because there is insufficient information to tell.

Expand

Correct

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? 1 / 1 point

- ○ The average performance of all their ornithologists (0.5%).
- ◉ The performance of their best ornithologist (0.3%).
- ○ The average of regular citizens of Peacetopia (1.2%).
- ○ The average of all the numbers above (0.66%).

Expand

Correct
Yes. The best human performance is closest to Bayes' error.

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

- ○ The learning algorithm's performance -> human-level performance -> Bayes error.
- ○ The learning algorithm's performance -> Bayes error -> human-level performance.
- ○ Human-level performance -> Bayes error -> the learning algorithm's performance.
- ◉ Human-level performance -> the learning algorithm's 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. You find that a team of ornithologists debating and discussing an image gets an even better 0.1% performance, so you define that as "human-level performance." After working further on your algorithm, you end up with the following: 1 / 1 point

Human-level performance	0.1%
Training set error	2.0%
Dev set error	2.1%

Based on the evidence you have, which two of the following four options seem the most promising to try? (Check two options.)

- ☑ Train a bigger model to try to do better on the training set.

  ✓ Correct

- ☑ Try decreasing regularization.

  ✓ Correct

- ☐ Try increasing regularization.
- ☐ Get a bigger training set to reduce 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.)

- [x] You have overfit to the dev set.

  ✓ Correct

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

  ✓ Correct

- [ ] You have underfitted to the dev set.

Expand

Correct
Great, you got all the right answers.

12. After working on this project for a year, you finally achieve: 1 / 1 point

Human-level performance	0.10%
Training set error	0.05%
Dev set error	0.05%

What can you conclude? (Check all that apply.)

- [ ] This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- [x] If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

  ✓ Correct

- [x] It is now harder to measure avoidable bias, thus progress will be slower going forward.

  ✓ Correct

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.
- ( ) 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.
- (•) Reset your "target" (metric) for the team and tune to it.

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 1 / 1 point 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?

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

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

- [ ] With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.
- [x] You could consider a tradeoff where you use a subset of the cat data to find reasonable performance with reasonable iteration pacing.

  ✓ **Correct**
  Yes. This is similar to satisficing metrics where "good enough" determines the size of the data.

- [x] Given a significant budget for cloud GPUs, you could mitigate the training time.

  ✓ **Correct**
  Yes. More resources will allow you to iterate faster.

- [x] Accuracy should exceed the City Council's requirements but the project may take as long as the bird detector because of the two week training/iteration time.

  ✓ **Correct**
  Yes. The 10x size increase adds a small amount of accuracy but takes too much time.

Expand

Correct
Great, you got all the right answers.