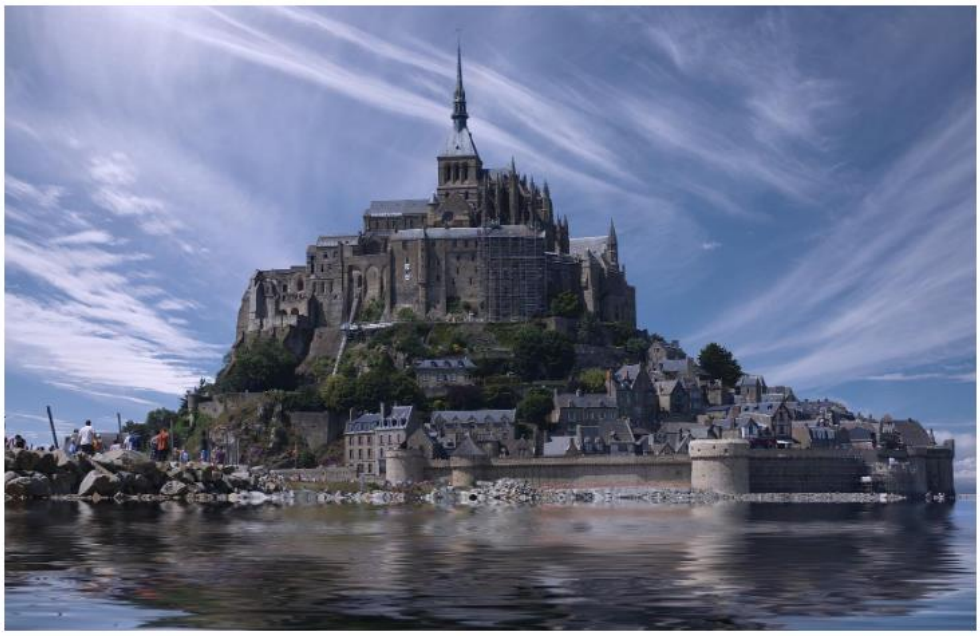
Congratulations! You passed!

Grade received	Latest Submission Grade	To pass
86.66%	86.67%	80% or higher

Go to next item

1. This example is adapted from a real production application, but with details disguised to protect confidentiality. 1 / 1 point



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 meet with them and ask for just one evaluation metric. True/False?

- ○ False
- ◉ True:

Expand

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

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
97%	1 sec	3MB

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

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. Which of the following best answers why it is important to identify optimizing and satisficing metrics? 1 / 1 point

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

Expand

Correct
Yes. Thresholds are essential for evaluation of key use case constraints.

4. With 10,000,000 data points, what is the best option for train/dev/test splits? 0 / 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

Incorrect
No. Given the size of the data, it isn't necessary to have so much data in the dev set to evaluate bias.

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

- (●) True
- ( ) False

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 proportionately to the train/dev/test sets. You object because: 1 / 1 point

- ○ The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- ◉ If we add the images to the test set then it won't reflect the distribution of data expected in production.
- ○ The additional data would significantly slow down training time.
- ○ The training set will not be as accurate because of the different distributions.

Expand

✓ Correct
Yes. Using the data in the training set could be beneficial, but you wouldn't want to include such images in your test set as they are not from the expected distribution of data you'll see in production.

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

- ○ Try an ensemble model to reduce bias and variance.
- ○ 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.

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

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.
- (●) True.

Expand

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

- [ ] Try increasing regularization.
- [x] Train a bigger model to try to do better on the training set.

Correct

- [x] Try decreasing regularization.

  ✓ Correct

- [ ] Get a bigger training set to reduce variance.

Expand

✓ Correct
Great, you got all the right 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) 1 / 1 point

- [ ] You have underfitted to the dev set.
- [ ] Try decreasing regularization for better generalization with the dev set.
- [x] You should try to get a bigger dev set.

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

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

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

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

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

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

- [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

- ( ) Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ( ) Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- (•) Rethink the appropriate metric for this task, and ask your team to tune to the new metric.
- ( ) Ask your team to take into account both accuracy and false negative rate during development.

Expand

Correct

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

- ○ Add pooling layers to downsample features to accommodate the new species.
- ○ 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.

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. (Wow Cat detectors are just incredibly useful, aren't they?) Because of years of working on Cat detectors, you have such a huge dataset of 100,000,000 cat images that training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.) 0 / 1 point

- ☑ Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.

  ✓ Correct

- ☐ Having built a good Bird detector, you should be able to take the same model and hyperparameters and just apply it to the Cat dataset, so there is no need to iterate.
- ☑ Needing two weeks to train will limit the speed at which you can iterate.

  ✓ Correct

- ☐ If 100,000,000 examples is enough to build a good enough Cat detector, you might be better off training with just 10,000,000 examples to gain a ≈10x improvement in how quickly you can run experiments, even if each model performs a bit worse because it's trained on less data.

Expand

✗ Incorrect
You didn't select all the correct answers