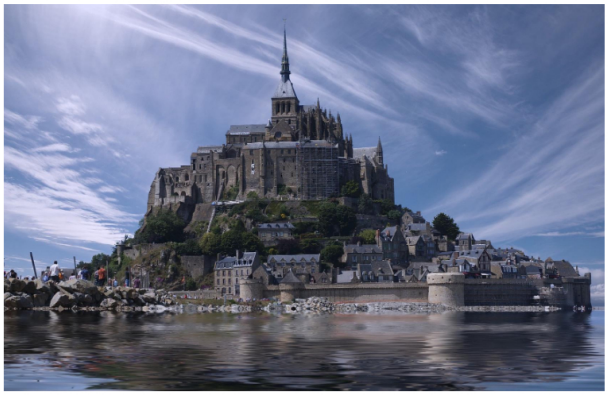
## Try again once you are ready

Grade received 73.33% | Latest Submission Grade 73.33% | To pass 80% or higher

Try again

**1. Problem Statement** 0 / 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

✗ **Incorrect**
No. The goal is to have one metric that focuses the development effort and increases iteration velocity.

**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?

- ◉ 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.
- ○ 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.

Expand

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

3. Which of the following best answers why it is important to identify optimizing and satisficing metrics? 0 / 1 point

- (•) Knowing the metrics provides input for efficient project planning.
- ( ) 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.

Expand

⊗ Incorrect
No. There may be some benefit to project planning but the most important value is in evaluation.

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 | 3,000,000 | 1,000,000 |

- ( )

| Train | Dev | Test |
|---|---|---|
| 3,333,334 | 3,333,334 | 3,333,334 |

- ( )

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

- (•)

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

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. The data can contribute to training the model.

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) 0 / 1 point

- [x] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.

! This should not be selected
No. This is a relatively small increase in data.

- [ ] 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.

- [ ] The dev set no longer reflects the distribution of data (security cameras) you most care about.

Expand

⊗ Incorrect
You didn't select all the correct 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

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

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 their volunteer amateur ornithologists.
- The performance of the average citizen of Peacetopia.
- The performance of the head of the City Council.

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**

- True.
- False.

Expand

**Correct**
Yes. By definition, human level error is worse than Bayes error.

**10.** Which of the following best expresses how to evaluate the next steps in your project when your results for human-level performance, train, and dev set error are 0.1%, 2.0%, and 2.1% respectively? **1 / 1 point**

- Evaluate the test set to determine the magnitude of the variance.
- Port the code to the target devices to evaluate if your model meets or exceeds the satisficing metrics.
- Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.
- Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.

Expand

**Correct**
Yes. Always choose the area with the biggest opportunity for improvement.

**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 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**

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

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

Expand

Correct
Great, you got all the right answers.

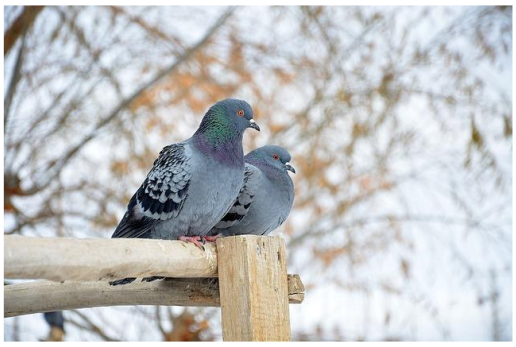
**13.** It turns out Peacetopia has hired one of your competitors to build a system as well. You and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! Still, when Peacetopia tries out both systems, they conclude they like 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.
- ( ) 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.
- (•) Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.

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 data is being tested on a new type of data. **1 / 1 point**



You have only 1,000 images of the new species of bird. The city expects a better system from you within the next 3 months. Which of these should you do first?

- ( ) Add the 1.000 images into your dataset and reshuffle into a new train/dev/test split.
- ( ) Try data augmentation/data synthesis to get more images of the new type of bird.
- ( ) Put the 1.000 images into the training set so as to try to do better on these birds.
- (•) Use the data you have to define a new evaluation metric (using a new dev/test set) taking into account the new species, and use that to drive further progress for your team.

Expand

Correct

**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 **1 / 1 point**

cat images that training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.)

- [ ] 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.
- [x] 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 $\approx$10x improvement in how quickly you can run experiments, even if each model performs a bit worse because it's trained on less data.

  ✓ Correct

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

  ✓ Correct

- [x] Needing two weeks to train will limit the speed at which you can iterate.

  ✓ Correct

Expand

**Correct**
Great, you got all the right answers.