Try again once you are ready

Grade received 66.66%
Latest Submission Grade 66.67%
To pass 80% or higher
Retake the assignment in **7h 44m**

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?

- (•) True:
- ( ) False

Expand

Correct
Yes. 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?

- Accuracy is an optimizing metric, therefore the most accurate model is the best choice.
- 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.
- Take the model with the smallest runtime because that will provide the most overhead to increase accuracy.

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? **1 / 1 point**

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

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? **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? **0 / 1 point**

- False
- True

Expand

Incorrect
No. Adding this data to the training set will change the training set distribution. However, it is not a

No. Adding this data to the training set will change the training set distribution. However, it is not a problem to have different training and dev distributions. In contrast, it would be very problematic to have different dev and test set distributions.

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 test set. You object because: **1 / 1 point**

- [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

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

  ✓ Correct

- [ ] The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- [ ] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.

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.
- ○ Yes, because having a 4.0% training error shows you have a high bias.
- ◉ No, because there is insufficient information to tell.
- ○ No, because this shows your variance is higher than your bias.

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 all the numbers above (0.66%).
- ○ The average of regular citizens of Peacetopia (1.2%).

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**

- ○ Human-level performance -> Bayes error -> the learning algorithm's performance.
- ○ The learning algorithm's performance -> human-level performance -> Bayes error.
- ○ The learning algorithm's performance -> Bayes error -> human-level 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: **0 / 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.)

- ☑ Get a bigger training set to reduce variance.

! **This should not be selected**

- ☐ Try decreasing regularization.
- ☑ Try increasing regularization.

! **This should not be selected**

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

Expand

Incorrect
You didn't select all the correct 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**

- ☐ Get a bigger test set to increase its accuracy.
- ☑ 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.
- ☑ Increase the size of the dev set.

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

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

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? **0 / 1 point**

- ( ) 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.
- ( ) 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.

Expand

Incorrect
No. This choice also points to the incorrect target.

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**

- (•) 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. (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

- [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 ≈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

- [ ] 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.

Expand

Incorrect
You didn't select all the correct answers