Congratulations! You passed!

Grade received 86.66% **Latest Submission Grade** 86.67% **To pass** 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?

- (●) 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?

- (•) 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.
- ( ) Create one metric by combining the three metrics and choose the best performing model.
- ( ) 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 metric types sets thresholds for satisficing metrics. This provides explicit evaluation criteria.
- ( ) 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 optimizing metric informs the team which models they should try first.

Expand

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

4. You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal? 1 / 1 point

- (•) With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.
- ( ) The emphasis on the training set will allow us to iterate faster.
- ( ) The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.
- ( ) The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.

Expand

Correct
Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

5. After setting up your train/dev/test sets, the City Council comes across another 1,000,000 images, called the “citizens’ data”. Apparently the citizens of Peacetopia are so scared of birds that they volunteered to take pictures of the sky and label them, thus contributing these additional 1,000,000 images. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. **1 / 1 point**

Notice that adding this additional data to the training set will make the distribution of the training set different from the distributions of the dev and test sets.

Is the following statement true or false?

"You should not add the citizens' data to the training set, because if the training distribution is different from the dev and test sets, then this will not allow the model to perform well on the test set."

- ○ True
- ◉ False

Expand

Correct
False is correct: Sometimes we'll need to train the model on the data that is available, and its distribution may not be the same as the data that will occur in production. Also, adding training data that differs from the dev set may still help the model improve performance on the dev set. What matters is that the dev and test set have the same distribution.

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.
- ☑ 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.

- ☑ 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.

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**

- ○ 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 you do not know what the human performance level is.

Expand

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

8. You ask a few people to label the dataset so as to find out what is human-level performance. You find the following levels of accuracy: 1 / 1 point

Bird watching expert #1	0.3% error
Bird watching expert #2	0.5% error
Normal person #1 (not a bird watching expert)	1.0% error
Normal person #2 (not a bird watching expert)	1.2% error

If your goal is to have "human-level performance" be a proxy (or estimate) for Bayes error, how would you define "human-level performance"?

- 0.4% (average of 0.3 and 0.5)
- 0.75% (average of all four numbers above)
- 0.3% (accuracy of expert #1)
- 0.0% (because it is impossible to do better than this)

Expand

Correct

9. Which of the below shows the optimal order of accuracy from worst to best? 0 / 1 point

- Human-level performance -> Bayes error -> the learning algorithm's performance.
- The learning algorithm's performance -> Bayes error -> human-level performance.
- Human-level performance -> the learning algorithm's performance -> Bayes error.
- The learning algorithm's performance -> human-level performance -> Bayes error.

Expand

Incorrect
No. HLP may be better than your algorithm's performance but it cannot be better than BE.

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.)

- [x] Try decreasing regularization.

Correct

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

Correct

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

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) **0 / 1 point**

- [ ] Try decreasing regularization for better generalization with the dev set.
- [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.

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

Expand

Incorrect
You didn't select all the correct 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 likely? (Check all that apply.) **1 / 1 point**

- [x] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.

Correct
Yes. Exceeding human performance means you are close to Bayes error.

- [ ] There is still avoidable bias.
- [x] The model has recognized emergent features that humans cannot. (Chess and Go for example)

Correct
Yes. When Google beat the world Go champion, it was recognized that it was making deeper moves than humans.

- [ ] This result is not possible since it should not be possible to surpass human-level performance.

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

- Ask your team to take into account both accuracy and false negative rate during 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.
- Pick false negative rate as the new metric, and use this new metric to drive all further 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.
- Split them between dev and test and re-tune.
- (•) Augment your data to increase the images of the new bird.
- Put the new species' images in training data to learn their features.

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

- [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.

- [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.

- [ ] With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.
- [x] Given a significant budget for cloud GPUs, you could mitigate the training time.

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

Expand

Correct
Great, you got all the right answers.