Your grade: 100%

Your level: **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 ease 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. The city revises its criteria to: 1 / 1 point

- **"It 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
- Find the subset of models that meet the runtime and memory criteria. Then, choose the highest accuracy.
- 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

- Knowing the metrics provides input for efficient project planning.
- 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.
- 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. You propose a 93.2%/6.8% 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

- The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.
- The emphasis on the training set will allow us to iterate faster.
- 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.
- With a dataset comprising 10M individual samples, 27K represents 270k samples, which is enough to evaluate the accuracy of the optimizing metric.

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 birds and upload them, bringing the total number of 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 (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 proportionately to the train/dev/test sets. You object because: 1 / 1 point

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

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. 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?

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

Expand

Correct

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

Expand

Correct
Yes. This is the peak of human performance in this task.

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

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. After working on your algorithm you have to decide the next steps. Currently, human-level performance is 0.2%, training is at 2.0% and the dev set is at 2.1%. Which, two of the following four, statements best describes your thought process? 1 / 1 point

- Address bias first through a larger model to get closer to human level error.

  Correct
  Yes. Selecting the largest difference from (train set error - human level error) and (dev set error - train set error) and reducing bias or variance accordingly is the most productive step.

- Get a bigger training set to reduce variance.
- Decrease variance via regularization on training and dev sets have similar performance.
- Decrease regularization to boost smaller signals.

  Correct
  Yes. Bias is higher than 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	0.2%
Dev set error	2.1%
Test set error	7.0%

What does this mean? (Check the two best options.)

- You have overfit to the dev set.

  Correct

- You should try to get a bigger dev set.

  Correct

- You have underfitted to the dev set.
- You should get a bigger test 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.
- You are close to Bayes error and possible overfitting.

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

- With only 0.05% further progress to make you should quickly be able to close the remaining gap to 0%
- An error of 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. 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

- Expand your model size to account for more corner cases.
- Look at the models you've developed during the development process and find the one with low false negative error. Use it.
- Reset your "target" metric for the team and tune to it.
- Pick false negative rate as the new metric, and use this new metric to drive all further development.

Expand

Correct
Yes. The target has shifted as 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 images of the new species. The city expects a better system from you within the next 3 months. Which of these ground you do first? 1 / 1 point

- Add pooling layers to downsample features to accommodate the new species.
- Pull the new species' images in training data to learn their features.
- Split them between dev and test and in 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. 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? 1 / 1 point

- Reducing the model complexity will allow the use of the larger data set but (presumed) accuracy.
- Lowering the number of images will reduce training time and likely allow for an acceptable tradeoff between iteration speed and accuracy.

  Correct
  Yes. This is a sweet spot that allows development at a reasonable rate without significant accuracy loss.

- This significantly impacts iteration speed.

  Correct
  Yes. The training time is an absolute constraint on iteration.

Expand

Correct
Great, you got all the right answers.