1.

Question 1

Problem Statement

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 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.

Note: Having three evaluation metrics makes it harder for you to quickly choose between two different algorithms, and will slow down the speed with which your team can iterate. True/False?

1 / 1 point

Expand

Correct

2.

Question 2

The city asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria?

1 / 1 point

Expand

Correct

Yes. The thresholds provide a way to evaluate models head to head.

3.

Question 3

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

0 / 1 point

Expand

Incorrect

No. The optimizing metric has no upper bound for performance.

4.

Question 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

Expand

Correct

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

5.

Question 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.

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."

1 / 1 point

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.

Question 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

Expand

Correct

Great, you got all the right answers.

7.

Question 7

You train a system, and its errors are as follows (error = 100%-Accuracy):

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?

1 / 1 point

Expand

Correct

8.

Question 8

You want to define what human-level performance is to the city council. Which of the following is the best answer?

0 / 1 point

Expand

Incorrect

No. The average reflects a range of skills, not the best.

9.

Question 9

Which of the below shows the optimal order of accuracy from worst to best?

0 / 1 point

Expand

Incorrect

No. in an optimal scenario, your algorithm's performance would be better than HLP but it can never be better than BE.

10.

Question 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:

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.)

1 / 1 point

Expand

Correct

Great, you got all the right answers.

11.

Question 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

Expand

Correct

Great, you got all the right answers.

12.

Question 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

Expand

Correct

Great, you got all the right answers.

13.

Question 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?

0 / 1 point

Expand

Incorrect

No. You must maintain accuracy and include false negatives.

14.

Question 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.

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?

1 / 1 point

Expand

Correct

15.

Question 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.)

1 / 1 point

Expand

Correct

Great, you got all the right answers.

1.

Question 1

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

1 / 1 point

Expand

Correct

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

2.

Question 2

The city asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria?

1 / 1 point

Expand

Correct

Yes. The thresholds provide a way to evaluate models head to head.

3.

Question 3

Based on the city's requests, which of the following would you say is true?

1 / 1 point

Expand

Correct

4.

Question 4

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

1 / 1 point

Expand

Correct

Yes. The size of the data set allows for bias and variance evaluation with smaller data sets.

5.

Question 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.

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."

1 / 1 point

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.

Question 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
Expand

Correct
Great, you got all the right answers.

7.
Question 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?

0 / 1 point
Expand

Incorrect
No. Unless you have strong reasons to believe the labeling is suspect, it's back to the drawing board.

8.
Question 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:

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"?

1 / 1 point
Expand

Correct

9.
Question 9
Which of the below shows the optimal order of accuracy from worst to best?

1 / 1 point
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.
Question 10

After working on your algorithm you have to decide the next steps. Currently, human-level performance is 0.1%, training is at 2.0% and the dev set is at 2.1%. Which statement below best describes your thought process?

0 / 1 point

Expand

Incorrect

You didn't select all the correct answers

11.

Question 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

Expand

Correct

Great, you got all the right answers.

12.

Question 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

Expand

Correct

Great, you got all the right answers.

13.

Question 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

Expand

Correct

Yes. The target has shifted so an updated metric is required.

14.

Question 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

Expand

Incorrect
No. Additional layers won't adjust the features in a meaningful way to learn the new species.

15.

Question 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
Expand

Correct
Great, you got all the right answers.