Quiz, 15 questions

### **\**

## **Congratulations! You passed!**

Next Item



1 / 1 points

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 labelled:

- 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 Peacetonia

security carrieras month reacceopia.

## Bird recognition in the city of Peacetopia (case study)

15/15 points (100%)

Quiz, 15 questions

- 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.

<u>Note</u>: 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?

0	True		
Corr	rect		
	False		

points

1/1

After further discussions, the city narrows down its criteria to:

- "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 10sec to classify a new image."
- "We want the model to fit in 10MB of memory."

If you had the three following models, which one would you choose?

Test Accuracy	Runtime	Memory size
97%	1 sec	3МВ

lest Accuracy Runtime Memory size		Test Accuracy	Runtime	Memory size
-----------------------------------	--	---------------	---------	-------------

99% 9MB 13 sec Bird recognition in the city of Peacetopia (case study) 15/15 points (100%) **Test Accuracy** Runtime Memory size Quiz, 15 questions 97% 3 sec 2MB **Test Accuracy** Runtime Memory size 98% 9 sec 9MB Correct Correct! As soon as the runtime is less than 10 seconds you're good. So, you may simply maximize the test accuracy after you made sure the runtime is <10sec. 1/1 points Based on the city's requests, which of the following would you say is true? Accuracy is an optimizing metric; running time and memory size are a satisficing metrics. Correct Accuracy is a satisficing metric; running time and memory size are an optimizing metric. Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three. Accuracy, running time and memory size are all satisficing metrics because you have to do sufficiently well on all three for your system to be acceptable.

**/** 

1/1 points

4.

## **Structuring your data**

Bird recognition implehencity of Peacet spine (caspistudy) ta into 15/15 points (100%)

Quiz, 15 questions train/dev/test sets. Which of these do you think is the best choice?

Train	Dev	Test
3,333,334	3,333,333	3,333,333

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

0	Train	Dev	Test
	9,500,000	250,000	250,000

#### Correct

Yes.

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



1/1 points

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.

You should not add the citizens' data to the training set, because this will cause the training and dev/test set distributions to become different, thus hurting dev and test set performance. True/False?

True False

Correct

## Adding this data to the training set will change the training set distribution. However, it is not a problem to have different Bird recognitioning the with the with the with the second the seco

Quiz, 15 questions

15/15 points (100%) problematic to have different dev and test set distributions.

1/1 points 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: A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set. **Un-selected is correct** The test set no longer reflects the distribution of data (security cameras) you most care about. Correct 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 The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data (similar to the New York City/Detroit housing prices example from lecture).

**Un-selected** is correct



1/1 points You train a system, and its errors are as follows (error = 100%-Accuracy):

Bird recogni	tion in the city of Peacetopia (ca	se <sub>4</sub> stµdy) <sub>15</sub>	/15 points (100%)
Quiz, 15 questions	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 having 4.0% training error shows you have high bias.
	Yes, because this shows your bias is higher than your variance.
	No, because this shows your variance is higher than your bias.
0	No, because there is insufficient information to tell.
Corre	ect



1/1 points

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"?

	0.0% (because it is impossible to do better than this)
0	0.3% (accuracy of expert #1)

### Correct

## Bird recognition in the city of Peacetopia (case study)

15/15 points (100%)

0.75% (average of all four numbers above)  1/1 points
ch of the following statements do you agree with?
A learning algorithm's performance can be better than human- level performance but it can never be better than Bayes error.
orrect
onecc
A learning algorithm's performance can never be better than human-level performance but it can be better than Bayes error.
A learning algorithm's performance can never be better than human-level performance nor better than Bayes error.
A learning algorithm's performance can be better than human- level performance and better than Bayes error.



1/1 points

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-

## Bird recognitionein the city of Pagentapia (6as gstudy) ou end up 5/15 points (100%)

Quiz, 15 questions

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.)

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

Correct

Try increasing regularization.

Un-selected is correct

Get a bigger training set to reduce variance.

Un-selected is correct

Try decreasing regularization.

Correct



1/1 points

11.

You also evaluate your model on the test set, and find the following:

Bird recogni Quiz, 15 questions	tion in the city of Peacetopia (case study)				
	Training set error	2.0%			
	Dev set error	2.1%			
	Test set error	7.0%			

15/15 points (100%)

What c	loes this mean? (Check the two best options.)			
	You have overfit to the dev set.			
Corr	ect			
	You should get a bigger test set.			
Un-selected is correct				
Corre	You should try to get a bigger dev set.			
	You have underfit to the dev set.			
Un-selected is correct				

12.

1/1 points

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%

What can you conclude? (Check all that apply.)



## This is a statistical anomaly (or must be the result of statistical

# noise) since it should not be possible to surpass human-level Bird recognition ipertheacity of Peacetopia (case study)

15/15 points (100%)

Quiz, 15 questions

	μ	1	•	<b>,</b>	13
Un-s	selected is correct				
	If the test set is big enough for accurate, this implies Bayes er				
Corr	ect				
	With only 0.09% further prograbe able to close the remaining			you should quickly	
Un-s	selected is correct				
	It is now harder to measure av	voidab	le bias	s, thus progress will	
Corre	ect				
system with al system and yo compe accura	1 / 1 points  s out Peacetopia has hired one on as well. Your system and your bout the same running time and has higher accuracy! However, our competitor's systems, they centior's system better, because ency, you have more false negative in the air). What should you do?	composited memons when concluded wen the conclusion conclusions are serifations.	etitor l lory size Peace de they lough	ooth deliver systems ze. However, your etopia tries out your / actually like your you have higher ove	rall
	Look at all the models you've of development process and find negative error rate.			_	
	Ask your team to take into acconegative rate during developm		ooth a	ccuracy and false	
0	Rethink the appropriate metri	c for tl	his tas	k, and ask your team	1

to tune to the new metric.

#### Correct

## Bird recognition in the city of Peacetopia (case study)

15/15 points (100%)

Quiz, 15 questions

Pick false negative rate as the new metric, and use this new metric to drive all further development.



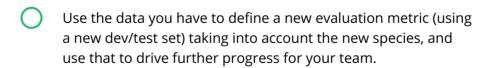
1/1 points

#### 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?



#### Correct

Put the 1,000 images into the training set so as to try to do better on these birds.

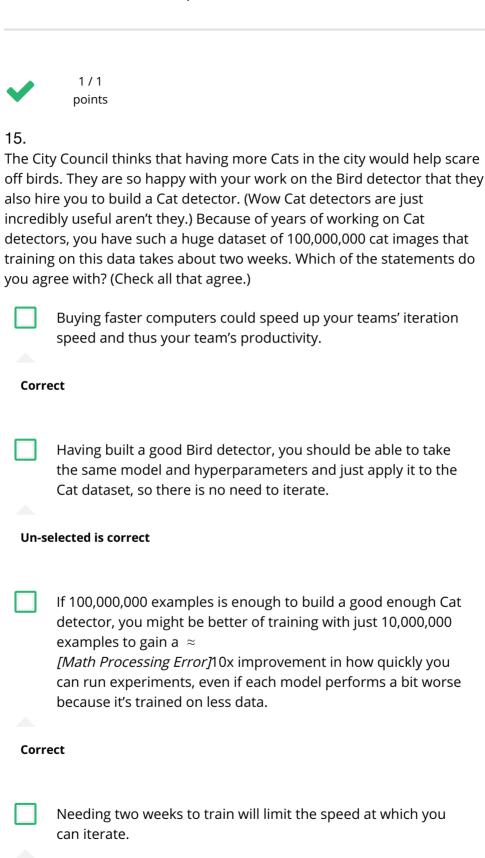
Try data augmentation/data synthesis to get more images of the new type of bird.

		CITC TICVV	type o	i bii a.			
D:1		: L1		- t D		/ <b>. . .</b> .	\
หาra	recognition	in the	CITW	OT PEA	ceronia	icase stiin	(7)
DILU	recognition	111 (110	. CILY	OI I Cu	cctopia	(case staa	y /

Quiz, 15 questions

Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.

15/15 points (100%)



Correct