Congratulations! You passed!

Grade received 86.66%
Latest Submission Grade 80%
To pass 80% or higher
Retake the assignment in **7h 44m**
Go to next item

1. This example is adapted from a real production application, but with details disguised to protect confidentiality. **0 / 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

Incorrect
No. More than one metric expands the choices and tradeoffs you have to decide for each with unknown effects on the other two.

2. After further discussions, the city narrows down 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."

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

○

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

○

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

○

Test Accuracy	Runtime	Memory size
97%	3 sec	2MB

◉

Test Accuracy	Runtime	Memory size

98%	9 sec	9MB

Expand

Correct
Correct! This model has the highest test accuracy, the prominent criteria you are looking for, compared with other models, and also has a runtime <10 seconds and memory size < 10MB.

3. Based on the city's requests, which of the following would you say is true? **1 / 1 point**

- (•) Accuracy is an optimizing metric; running time and memory size are satisfying metrics.
- ( ) Accuracy is a satisfying metric; running time and memory size are an optimizing metric.
- ( ) Accuracy, running time and memory size are all satisfying metrics because you have to do sufficiently well on all three for your system to be acceptable.
- ( ) Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.

Expand

Correct

4. Structuring your data **1 / 1 point**

Before implementing your algorithm, you need to split your data into train/dev/test sets. Which of these do you think is the best choice?

( )

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

(•)

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

( )

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

( )

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

Expand

Correct
Yes.

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. Which of the following is the best use of that additional data? **0 / 1 point**

- ( ) Split it among train/dev/test equally.
- ( ) Add it to the training set.
- (•) Do not use the data. It will change the distribution of any set it is added to.
- ( ) Add it to the dev set to evaluate how well the model generalizes across a broader set.

Expand

Incorrect
No. The data can contribute to training the model.

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

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. 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? **1 / 1 point**

- (•) Train a bigger network to drive down the >4.0% training error.
- ( ) Get more data or apply regularization to reduce variance.
- ( ) Validate the human data set with a sample of your data to ensure the images are of sufficient quality.
- ( ) Try an ensemble model to reduce bias and variance.

Expand

Correct
Yes. Avoidable bias is >4.2% which is larger than the 2.1% variance.

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"?

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

Expand

Correct

9. A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error. True/False? **0 / 1 point**

- ( ) True.
- (•) False.

Expand

Incorrect
No. By definition, human level error is worse than Bayes error.

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, two of the following four, statements best describe your thought process? **1 / 1 point**

- [ ] Decrease variance via regularization so training and dev sets have similar performance.
- [x] Decrease regularization to boost smaller signals.

Correct
Yes. Bias is higher than variance.

- [x] Address bias first through a larger model to get closest to human level error.

Correct
Yes. Selecting the largest difference from (train set error - human level error) and (dev

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.

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	0.1%
Training set error	2.0%
Dev set error	2.1%
Test set error	7.0%

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

- [x] You have overfit to the dev set.

Correct

- [ ] You should get a bigger test set.
- [ ] You have underfitted to the dev set.
- [x] You should try to get a bigger dev set.

Correct

Expand

Correct
Great, you got all the right answers.

12. After working on this project for a year, you finally achieve: **1 / 1 point**

Human-level performance	0.10%
Training set error	0.05%
Dev set error	0.05%

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

- [x] If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

Correct

- [ ] 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] It is now harder to measure avoidable bias, thus progress will be slower going forward.

Correct

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? **1 / 1 point**

- ( ) Apply regularization to minimize the false negative rate.
- ( ) 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.
- ( ) 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 so an updated metric is required.

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? **1 / 1 point**

- Add the new images and split them among train/dev/test.
- Add hidden layers to further refine feature development.
- Put them into the dev set to evaluate the bias and re-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. (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**

- [ ] 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.
- [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 $\approx$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] Needing two weeks to train will limit the speed at which you can iterate.

  Correct
- [x] Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.

  Correct

Expand

Correct
Great, you got all the right answers.