Back **Bird Recognition in the City of Peacetopia (Case Study)** coursera
Graded Quiz • 45 min
Due Apr 11, 2:59 PM CST

Congratulations! You passed!

Grade received 90% **To pass** 80% or higher

Bird Recognition in the City of Peacetopia (Case Study)

Quiz • 45 min

Bird Recognition in the City of Peacetopia (Case Study)

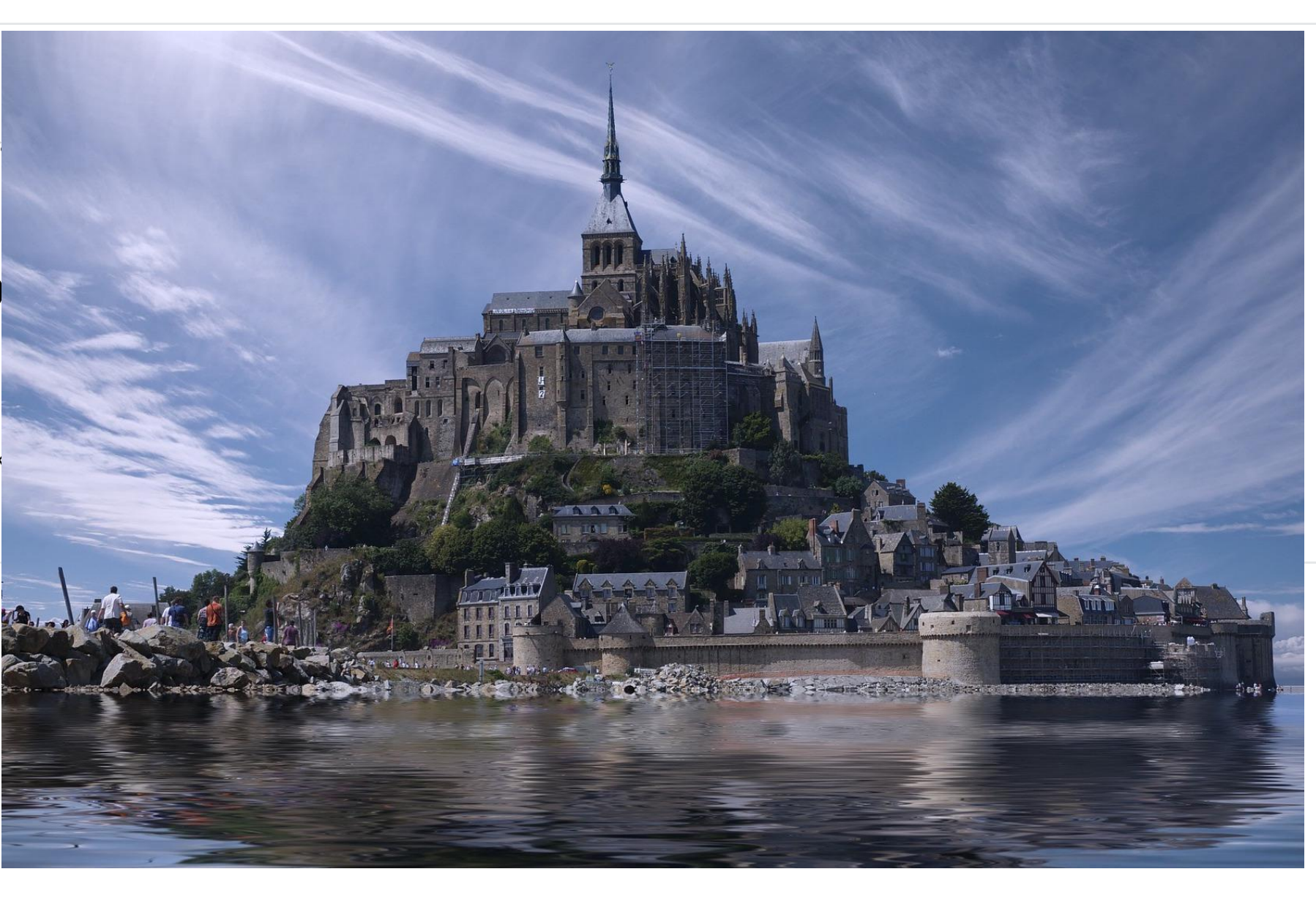
Latest Submission Grade 90%

Due Apr 11, 2:59 PM CST **Attempts** 3 every 8 hours

1. Problem Statement **1 / 1 point**

Try again

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?

(•) True

( ) False

Correct

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%	3 sec	2MB

( )

Test Accuracy	Runtime	Memory size
97%	1 sec	3MB

( )

Test Accuracy	Runtime	Memory size
99%	13 sec	9MB

(•)

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 make sure the runtime is <10 seconds.

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 metric types sets thresholds for satisficing metrics. This provides explicit evaluation criteria.

( ) 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.

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

4. With 10,000,000 data points, what is the best option for train/dev/test splits? **1 / 1 point**

(•) train - 95%, dev - 2.5%, test - 2.5%

( ) train - 60%, dev - 10%, test - 30%

( ) train - 33.3%, dev - 33.3%, test - 33.3%

( ) train - 60%, dev - 30%, test - 10%

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

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? **1 / 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.

Correct
Yes. It is not a problem to have different training and dev distributions. Different dev and test distributions would be an issue.

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**

(•) If we add the images to the test set then it won't reflect the distribution of data expected in production.

( ) The training set will not be as accurate because of the different distributions.

( ) The additional data would significantly slow down training time.

( ) 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 the lecture).

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? **0 / 1 point**

( ) Train a bigger network to drive down the >4.0% training error.

( ) Validate the human data set with a sample of your data to ensure the images are of sufficient quality.

( ) Get more data or apply regularization to reduce variance.

(•) Try an ensemble model to reduce bias and variance.

Incorrect
No. A best practice is to address the largest gap first.

8. You want to define what human-level performance is to the city council. Which of the following is the best answer? **1 / 1 point**

( ) The average of regular citizens of Peacetopia (1.2%).

( ) The average of all the numbers above (0.66%).

( ) The average performance of all their ornithologists (0.5%).

(•) The performance of their best ornithologist (0.3%).

Correct
Yes. The best human performance is closest to Bayes' error.

9. Which of the following statements do you agree with? **1 / 1 point**

( ) 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.

(•) A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error.

( ) A learning algorithm's performance can never be better than human-level performance but it can be better than Bayes error.

Correct

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? **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.

- [ ] Get a bigger training set to reduce 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 set error - train set error) and reducing bias or variance accordingly is the most productive step.

11. After running your model with the test set you find it is a 7.0% error compared to a 2.1% error for the dev set and 2.0% for the training set. What can you conclude? (Choose all that apply) **1 / 1 point**

- [x] You should try to get a bigger dev set.

Correct
Yes. The dev set performance versus the test set indicates it is overfitting.

- [ ] You have underfitted to the dev set.
- [ ] Try decreasing regularization for better generalization with the dev set.
- [x] You have overfitted to the dev set.

Correct
Yes. The dev set performance versus the test set indicates it is overfitting.

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**

- [ ] With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- [ ] This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- [x] All or almost all of the avoidable bias has been accounted for.

Correct
Yes. Exceeding human performance makes the identification of avoidable bias very challenging.

- [x] You are close to Bayes error and possible overfitting.

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

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**

( ) Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.

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

(•) Reset your "target" (metric) for the team and tune to it.

( ) Expand your model size to account for more corner cases.

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. Which of these should you do first? **1 / 1 point**

(•) Augment your data to increase the images of the new bird.

( ) Add hidden layers to further refine feature development.

( ) Add the new images and split them among train/dev/test.

( ) Put them into the dev set to evaluate the bias and re-tune.

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.) **0.5 / 1 point**

- [ ] 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.
- [x] Given a significant budget for cloud GPUs, you could mitigate the training time.

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

- [x] With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.

This should not be selected
No. Although you may have gained many insights that may reduce the number of iterations needed, a DeepLearning model requires multiple iterations when working in a new dataset.

- [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.