- **Introduction to ML Strategy**
- **Connect with your Mentors and Fellow Learners on Discourse!**
- **Setting Up your Goal**
- **Comparing to Human-level Performance**
- **Lecture Notes (Optional)**
- **Machine Learning Flight Simulator**
  - **Reading:** ReadingMachine Learning Flight Simulator

    . Duration: 2 minutes2 min
  - **Quiz:** Bird Recognition in the City of Peacetopia (Case Study)

    15 questions
- **Heroes of Deep Learning (Optional)**

Bird Recognition in the City of Peacetopia (Case Study)

Quiz1 hour 15 minutes • 1h 15m

Submit your assignment

Due May 21, 11:59 PM WIBMay 21, 11:59 PM WIB
Attempts 3 every 24 hours

Try again

Receive grade

To Pass 80% or higher

Your grade

73.33%

View Feedback

We keep your highest score

Like
Dislike
Report an issue

Back

Bird Recognition in the City of Peacetopia (Case Study)

Graded Quiz. • 1h 15m

DueMay 21, 11:59 PM WIB

Try again once you are ready

Grade received 73.33%
Latest Submission Grade 73.33%
To pass 80% or higher
Try again

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?

- (•) 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 revises 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 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?

- (•) 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.
- ( ) Accuracy is an optimizing metric, therefore the most accurate model is the best choice.
- ( ) Take the model with the smallest runtime because that will provide the most overhead to increase accuracy.

1 / 1 point

Expand

Correct

Yes. Once you meet the runtime and memory thresholds, accuracy should be maximized.

3.

Question 3

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

- ( ) 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 is an optimizing metric; running time and memory size are satisfying metrics.
- ( ) Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.

1 / 1 point

Expand

Correct

4.

Question 4

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

- ( ) train - 33.3%, dev - 33.3%, test - 33.3%
- (•) train - 95%, dev - 2.5%, test - 2.5%
- ( ) train - 60%, dev - 10%, test - 30%
- ( )

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

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."

- True
- False

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 dev set. You object because: (Choose all that apply)

- [x] 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

Yes. Adding a different distribution to the dev set will skew bias.

- [ ] The 1,000,000 citizens’ data images do not have a consistent x-->y mapping as the rest of the data.
- [ ] A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.

☑ The dev set no longer reflects the distribution of data (security cameras) you most care about.

Correct

Yes. The performance of the model should be evaluated on the same distribution of images it will see in production.

1 / 1 point

Expand

Correct

Great, you got all the right answers.

7.

Question 7

You train a system, and the train/dev set errors are 3.5% and 4.0% respectively. You decide to try regularization to close the train/dev accuracy gap. Do you agree?

- Yes, because having a 4.0% training error shows you have a high bias.
- No, because you do not know what the human performance level is.
- Yes, because this shows your bias is higher than your variance.
- ◉ No, because this shows your variance is higher than your bias.

0 / 1 point

Expand

Incorrect

No. Test accuracy is not given so we can't speak about variance.

8.

Question 8

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

- The average performance of all their ornithologists (0.5%).
- The average of all the numbers above (0.66%).
- The average of regular citizens of Peacetopia (1.2%).
- ◉ The performance of their best ornithologist (0.3%).

1 / 1 point

Expand

Correct

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

9.

Question 9
Which of the following statements do you agree with?

- 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 be better than human-level performance but it can never be better than Bayes error.
- (selected) A learning algorithm's performance can be better than human-level performance and better than Bayes error.
- A learning algorithm's performance can never be better than human-level performance nor better than Bayes error.

0 / 1 point

Expand

Incorrect

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.)

- [x] Try decreasing regularization.

Correct

- [ ] Get a bigger training set to reduce variance.
- [x] Train a bigger model to try to do better on the training set.

Correct

- [ ] Try increasing regularization.

1 / 1 point

Expand

Correct

Great, you got all the right answers.

11.

Question 11
You also evaluate your model on the test set, and find the following:

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

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 likely? (Check all that apply.)

- [ ] There is still avoidable bias.

This should not be selected

No. Exceeding human performance makes the identification of avoidable bias very challenging.

- [x] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.

Correct

Yes. Exceeding human performance means you are close to Bayes error.

- [ ] This result is not possible since it should not be possible to surpass human-level performance.
- [x] The model has recognized emergent features that humans cannot. (Chess and Go for example)

Correct

Yes. When Google beat the world Go champion, it was recognized that it was making deeper moves than humans.

0 / 1 point

Expand

Incorrect
You chose the extra incorrect 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?

- (●) 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.
- ( ) Look at all the models you’ve developed during the development process and find the one with the lowest false negative error rate.
- ( ) Expand your model size to account for more corner cases.

0 / 1 point
Expand

Incorrect
No. This choice also points to the incorrect target.

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?

- Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.
- Try data augmentation/data synthesis to get more images of the new type of bird.
- Put the 1,000 images into the training set so as to try to do better on these birds.
- Use the data you have to define a new evaluation metric (using a new dev/test set) taking into account the new species, and use that to drive further progress for your team.

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.)

- [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 ≈≈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

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

Correct

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

Correct

1 / 1 point

Expand

Correct

Great, you got all the right answers.