Congratulations! You passed!

Grade received 93.33% **Latest Submission Grade** 93.33% **To pass** 80% or higher

Go to next item

1. Problem Statement 1 / 1 point

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

Expand

Correct

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**

- Suggest to them that they focus on whichever criterion is to be optimized and then eliminate the other two.
- Suggest to them that they define which criterion is to be optimized. Then, set thresholds for the other two.
- Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.

Expand

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

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

- Accuracy is a satisfying metric; running time and memory size are an optimizing metric.
- 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.
- 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.

Expand

Correct

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

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

Expand

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

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. **1 / 1 point**

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

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. 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**

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

  ✓ **Correct**

- ☐ A bigger test set will slow down the speed of iterating because of the computational expense of evaluating models on the test set.
- ☑ 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.

Expand

Correct
Great, you got all the right answers.

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? **1 / 1 point**

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

Expand

Correct
Yes. You need to know what the human performance level is to estimate avoidable bias.

8. 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**

- The performance of the head of the City Council.
- The performance of the average citizen of Peacetopia.
- (•) The best performance of a specialist (ornithologist) or possibly a group of specialists.
- The performance of their volunteer amateur ornithologists.

Expand

Correct
Yes. This is the peak of human performance in this task.

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

- (•) 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 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 never be better than human-level performance but it can be better than Bayes error.

Expand

Correct

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: **1 / 1 point**

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

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

Correct

- [ ] Try increasing regularization.
- [ ] Get a bigger training set to reduce variance.

Expand

Correct
Great, you got all the right answers.

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 have overfitted to the dev set.

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

- [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.

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.)

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

Correct

- [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. 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**

- Pick false negative rate as the new metric, and use this new metric to drive all further development.
- Expand your model size to account for more corner cases.
- 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

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 data is being tested on a new type of data. **0 / 1 point**

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?

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

Expand

Incorrect
The true data distribution is changed. It means you need to adjust your evaluation. Because you evaluate your learning algorithm on dev and test sets, adding more data only to the training set doesn't help the algorithm to perform better.

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**

- [ ] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.
- [x] Lowering the number of images will reduce training time and likely allow for an acceptable tradeoff between iteration speed and accuracy.

Correct

Yes. There is a sweet spot that allows development at a reasonable rate without significant accuracy loss.

- [x] This significantly impacts iteration speed.

Correct

Yes. This training time is an absolute constraint on iteration.

Expand

Correct

Great, you got all the right answers.