Problem Statement

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 the following that they want an algorithm that

. Has high accuracy.

. Runs quickly and takes only a short time to classify a new image.

. 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 are delighted because this list of criteria will speed development and provide guidance on how to evaluate two different algorithms. True/False?

- True:
- False

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 point

- Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.
- 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.

3.The essential difference between an optimizing metric and satisficing metrics is the priority assigned by the stakeholders. True/False? 1 point

- False
- True

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

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

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. You should add the citizens' data to the training set. True/False? 1 point

- False
- True

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 point

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

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

Training set error	4.0%
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 this shows your bias is higher than your variance.
- Yes, because having a 4.0% training error shows you have a high bias.
- No, because this shows your variance is higher than your bias.

- ◯ No, because there is insufficient information to tell.

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 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”?

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

9. Which of the below shows the optimal order of accuracy from worst to best? 1 point

- ◯ The learning algorithm’s performance -> Bayes error -> human-level performance.
- ◯ Human-level performance -> the learning algorithm’s performance -> Bayes error.
- ◯ The learning algorithm’s performance -> human-level performance -> Bayes error.
- ◯ Human-level performance -> Bayes error -> the learning algorithm’s performance.

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 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.)

- ☐ Train a bigger model to try to do better on the training set.
- ☐ Try increasing regularization.
- ☐ Try decreasing regularization.
- ☐ Get a bigger training set to reduce variance.

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

- ☐ You should try to get a bigger dev set.
- ☐ You have overfit to the dev set.
- ☐ You have underfitted to the dev set.

- [ ] You should get a bigger test set.

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.) 1 point

- [ ] There is still avoidable bias.
- [ ] Pushing to even higher accuracy will be slow because you will not be able to easily identify sources of bias.
- [ ] This result is not possible since it should not be possible to surpass human-level performance.
- [ ] The model has recognized emergent features that humans cannot. (Chess and Go for example)

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 point

- ◯ Expand your model size to account for more corner cases.
- ◯ Reset your “target” (metric) for the team and tune to it.
- ◯ Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ◯ Look at all the models you’ve developed during the development process and find the one with the lowest false negative error rate.

14. 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. 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 point

- ◯ Add pooling layers to downsample features to accommodate the new species.
- ◯ Augment your data to increase the images of the new bird.
- ◯ Split them between dev and test and re-tune.
- ◯ Put the new species’ images in training data to learn their features.

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 point

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