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?

- ○ False
- ◉ True:

Expand

Correct
Yes. The goal is to have one metric that focuses the development effort and increases iteration velocity.

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?

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

Expand

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

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

- ◉ False
- ○ True

Expand

Correct
Yes. Satisficing metrics have thresholds for measurement and an optimizing metric is unbounded.

4. You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal? 1/

- ○ The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.
- ○ The emphasis on the training set will allow us to iterate faster.
- ○ The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.
- ◉ With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.

Expand

Correct
Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

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

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:

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

  ✓ Correct

- ☐ 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.
- ☑ 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

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?

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

Expand

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

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:

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)
- ○ 0.4% (average of 0.3 and 0.5)
- ◉ 0.3% (accuracy of expert #1)
- ○ 0.75% (average of all four numbers above)

Which of the below shows the optimal order of accuracy from worst to best?

- ○ The learning algorithm's performance -> human-level performance -> Bayes error.
- ○ The learning algorithm's performance -> Bayes error -> human-level performance.
- ○ Human-level performance -> Bayes error -> the learning algorithm's performance.
- ◉ Human-level performance -> the learning algorithm's performance -> Bayes error.

Expand

✓ **Correct**
Yes. A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error.

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.)

- ☐ Get a bigger training set to reduce variance.
- ☑ Train a bigger model to try to do better on the training set.

✓ **Correct**

- ☑ Try decreasing regularization.

✓ **Correct**

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)

- ☑ You have overfitted to the dev set.

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

- ☐ Try decreasing regularization for better generalization with the dev set.
- ☑ 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.

2. 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.
- [ ] The model has recognized emergent features that humans cannot. (Chess and Go for example)
- [x] This result is not possible since it should not be possible to surpass human-level performance.

! **This should not be selected**
No. It is possible to exceed human performance.

- [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.

. 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.)

- [x] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.

! **This should not be selected**
No. Fewer layers could result in lower accuracy that is not offset by the lower training time.

- [ ] This significantly impacts iteration speed.
- [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.

. 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 / 1 po

- ○ False
- ◉ True

Expand

✓ **Correct**
Yes. This will cause the training and dev/test set distributions to become different, however as long as dev/test distributions are the same you are aiming at the same target.

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:

- [ ] The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- [x] The test set no longer reflects the distribution of data (security cameras) you most care about.

✓ Correct

- [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

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

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 this shows your bias is higher than your variance.
- ( ) No, because this shows your variance is higher than your bias.
- ( ) 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.

Expand

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

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.

Expand

✓ Correct
Yes. The target has shifted so an updated metric is required.

5. 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 p

- [ ] Reducing the model complexity will allow the use of the larger data set but preserve accuracy.
- [x] This significantly impacts iteration speed.

  ✓ **Correct**
  Yes. This training time is an absolute constraint on iteration.

- [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.