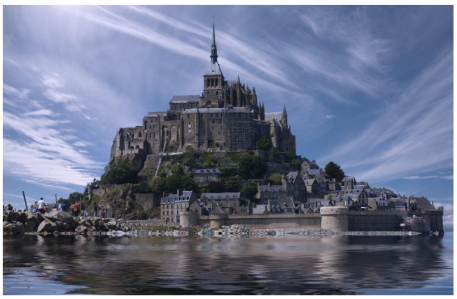
## 축하합니다! 통과하셨습니다!

받은 학점 100% 최신 제출물 학점 100% 통과 점수: 80% 이상

다음 항목으로 이동

**1. Problem Statement** 1 / 1점

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

더 보기

맞습니다

**2.** The city revises its criteria to: 1 / 1점

- "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?

- ( ) Create one metric by combining the three metrics and choose the best performing model.
- (●) Find the subset of models that meet the runtime and memory criteria. Then, choose the highest accuracy.
- ( ) Take the model with the smallest runtime because that will provide the most overhead to increase accuracy.
- ( ) Accuracy is an optimizing metric, therefore the most accurate model is the best choice.

더 보기

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

**3.** Which of the following best answers why it is important to identify optimizing and satisficing metrics? 1 / 1점

- (●) 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.
- ( ) Knowing the metrics provides input for efficient project planning.

더 보기

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

**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 / 1점

- ( ) The emphasis on the training set will allow us to iterate faster.
- ( ) 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 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.

더 보기

맞습니다
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. 1 / 1점

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

더 보기

맞습니다
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 / 1점

- ☑ 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.
- ☐ The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- ☑ 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

더 보기

맞습니다
Great, you got all the right answers.

**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? 1 / 1점

- ○ Get more data or apply regularization to reduce variance.
- ○ Try an ensemble model to reduce bias and variance.
- ○ 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.

더 보기

맞습니다
Yes. Avoidable bias is >4.2% which is larger than the 2.1% variance.

**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 / 1점

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

더 보기

맞습니다

**9.** A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error. True/False? 1 / 1점

- ○ False.
- ◉ True.

더 보기

맞습니다
Yes. By definition, human level error is worse than Bayes error.

**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 / 1점

- ☑ Decrease regularization to boost smaller signals.

✓ Correct
Yes. Bias is higher than variance.

- ☐ Decrease variance via regularization so training and dev sets have similar performance.

- [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.

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

더 보기

맞습니다
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 / 1점

- [ ] 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.

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

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

더 보기

맞습니다
Great, you got all the right answers.

**12.** After working on this project for a year, you finally achieve: 1 / 1점

| | |
|---|---|
| Human-level performance | 0.10% |
| Training set error | 0.05% |
| Dev set error | 0.05% |

What can you conclude? (Check all that apply.)

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

  ✓ **Correct**

- [ ] 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] It is now harder to measure avoidable bias, thus progress will be slower going forward.

  ✓ **Correct**

더 보기

맞습니다
Great, you got all the right answers.

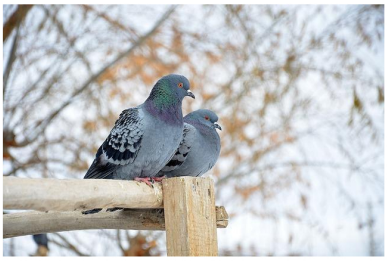
**13.** Your system is now very accurate but has a higher false negative rate than the City Council of Peacetopia would like. What is your best next step? 1 / 1점

- ( ) Expand your model size to account for more corner cases.
- ( ) 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.
- (•) Reset your "target" (metric) for the team and tune to it.

더 보기

맞습니다
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. 1 / 1점



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?

- ( ) 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.
- ( ) Put the 1,000 images into the training set so as to try to do better on these birds.

더 보기

맞습니다

**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 / 1점

- [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] 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.

  ✓ **Correct**
  Yes. The 10x size increase adds a small amount of accuracy but takes too much time.

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

더 보기

**맞습니다**
Great, you got all the right answers.