## 축하합니다! 통과하셨습니다!

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

23h 49m 후에 과제를 다시 풀어보세요.

다음 항목으로 이동

**1.** To help you practice strategies for machine learning, this week we'll present another scenario and ask how you would act. We think this "simulator" of working in a machine learning project will give an idea of what leading a machine learning project could be like!

1 / 1점

You are employed by a startup building self-driving cars. You are in charge of detecting road signs (stop sign, pedestrian crossing sign, construction ahead sign) and traffic signals (red and green lights) in images. The goal is to recognize which of these objects appear in each image. As an example, this image contains a pedestrian crossing sign and red traffic lights.

$$y^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{matrix} \text{"stop sign"} \\ \text{"pedestrian crossing sign"} \\ \text{"construction ahead sign"} \\ \text{"red traffic light"} \\ \text{"green traffic light"} \end{matrix}$$

Your 100,000 labeled images are taken using the front-facing camera of your car. This is also the distribution of data you care most about doing well on. You think you might be able to get a much larger dataset off the internet, which could be helpful for training even if the distribution of internet data is not the same.

You are getting started with this project. What is the first thing you do? Assume each of the steps below would take about an equal amount of time (a few days).

- Invest a few days in thinking on potential difficulties, and then some more days brainstorming about possible solutions. before training any model.
- Spend some time searching the internet for the data most similar to the conditions you expect on production.
- Spend a few days collecting more data using the front-facing camera of your car. to better understand how much data per unit time you can collect.
- **Train a basic model and do error analysis.**

더 보기

맞습니다
Applied ML is highly iterative. Having a basic model to do an error analysis can point you in the most promising directions with a lot of certainties.

**2.** Your goal is to detect road signs (stop sign, pedestrian crossing sign, construction ahead sign) and traffic signals (red and green lights) in images. The goal is to recognize which of these objects appear in each image. You plan to use a deep neural network with ReLU units in the hidden layers. For the output layer, which of the following gives you the most appropriate activation function?

1 / 1점

- Softmax
- ReLU
- Linear
- **Sigmoid**

더 보기

맞습니다
Correct. This works well since the output would be valued between 0 and 1 which represents the probability that one of the possibilities is present in an image.

**3.** You are carrying out error analysis and counting up what errors the algorithm makes. Which of these datasets do you think you should manually go through and carefully examine, one image at a time?

1 / 1점

- 500 randomly chosen images
- **500 images on which the algorithm made a mistake**
- 10,000 images on which the algorithm made a mistake
- 10,000 randomly chosen images

더 보기

맞습니다
Focus on images that the algorithm got wrong. Also, 500 is enough to give you a good initial sense of the error statistics. There's probably no need to look at 10,000, which will take a long time.

**4.** After working on the data for several weeks, your team ends up with the following data:

1 / 1점

- 100,000 labeled images taken using the front-facing camera of your car.
- 900,000 labeled images of roads downloaded from the internet.
- Each image's labels precisely indicate the presence of any specific road signs and traffic signals or combinations of them. For example, $y^{(i)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ means the image contains a stop sign and a red traffic light.

Because this is a multi-task learning problem, you need to have all your $y^{(i)}$ vectors fully labeled. If one example is equal to $\begin{bmatrix} 0 \\ ? \\ 1 \\ 1 \\ ? \end{bmatrix}$ then the learning algorithm will not be able to use that example. True/False?

- True
- **False**

더 보기

맞습니다
As seen in the lecture on multi-task learning, you can compute the cost such that it is not influenced by the fact that some entries haven't been labeled.

**5.** The distribution of data you care about contains images from your car's front-facing camera; which comes from a

1 / 1점

different distribution than the images you were able to find and download off the internet. How should you split the dataset into train/dev/test sets?

- Mix all the 100,000 images with the 900,000 images you found online. Shuffle everything. Split the 1,000,000 images dataset into 600,000 for the training set, 200,000 for the dev set and 200,000 for the test set.
- Mix all the 100,000 images with the 900,000 images you found online. Shuffle everything. Split the 1,000,000 images dataset into 980,000 for the training set, 10,000 for the dev set and 10,000 for the test set.
- Choose the training set to be the 900,000 images from the internet along with 20,000 images from your car's front-facing camera. The 80,000 remaining images will be split equally in dev and test sets.
- (선택됨) Choose the training set to be the 900,000 images from the internet along with 80,000 images from your car's front-facing camera. The 20,000 remaining images will be split equally in dev and test sets.

더 보기

맞습니다
Yes. As seen in the lecture, it is important that your dev and test set have the closest possible distribution to "real" data. It is also important for the training set to contain enough "real" data to avoid having a data-mismatch problem.

**6.** Assume you've finally chosen the following split between the data: 1 / 1점

| Dataset: | Contains: | Error of the algorithm: |
|---|---|---|
| Training | 940,000 images randomly picked from (900,000 internet images + 60,000 car's front-facing camera images) | 1% |
| Training-Dev | 20,000 images randomly picked from (900,000 internet images + 60,000 car's front-facing camera images) | 5.1% |
| Dev | 20,000 images from your car's front-facing camera | 5.6% |
| Test | 20,000 images from the car's front-facing camera | 6.8% |

You also know that human-level error on the road sign and traffic signals classification task is around 0.5%. Which of the following is true?

- You have a high bias.
- You have a large data-mismatch problem.
- The size of the train-dev set is too high.
- (선택됨) You have a high variance problem.

더 보기

맞습니다
Correct. Since the difference between the training-dev error and the training error is high.

**7.** Assume you've finally chosen the following split between the data: 1 / 1점

| Dataset: | Contains: | Error of the algorithm: |
|---|---|---|
| Training | 940,000 images randomly picked from (900,000 internet images + 60,000 car's front-facing camera images) | 8.8% |
| Training-Dev | 20,000 images randomly picked from (900,000 internet images + 60,000 car's front-facing camera images) | 9.1% |
| Dev | 20,000 images from your car's front-facing camera | 14.3% |
| Test | 20,000 images from the car's front-facing camera | 14.8% |

You also know that human-level error on the road sign and traffic signals classification task is around 0.5%. Based on the information given, a friend thinks that the training data distribution is much easier than the dev/test distribution. What do you think?

- Your friend is right. (I.e., Bayes error for the training data distribution is probably lower than for the dev/test distribution.)
- Your friend is wrong. (I.e., Bayes error for the training data distribution is probably higher than for the dev/test distribution.)
- (선택됨) There's insufficient information to tell if your friend is right or wrong.

더 보기

맞습니다
The algorithm does better on the distribution of data it trained on. But you don't know if it's because it trained on that distribution or if it really is easier. To get a better sense, measure human-level error separately on both distributions.

**8.** You decide to focus on the dev set and check by hand what the errors are due to. Here is a table summarizing your discoveries: 1 / 1점

| | |
|---|---|
| Overall dev set error | 15.3% |
| Errors due to incorrectly labeled data | 4.1% |
| Errors due to foggy pictures | 3.0% |
| Errors due to partially occluded elements. | 7.2% |
| Errors due to other causes | 1.0% |

In this table, 4.1%, 7.2%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about 7.2/15.3 = 47% of your errors are due to partially occluded elements.

You shouldn't invest all your efforts to get more images with partially occluded elements since 4.1 + 3.0 + 1.0 = 8.1 > 7.2. True/False?

- (선택됨) False
- True

더 보기

맞습니다
Correct. These kinds of arguments don't help us to decide on the strategy to follow. Other factors should be used, such as the tradeoff between the cost of getting new images and the improvement of the system performance.

**9.** You decide to focus on the dev set and check by hand what the errors are due to. Here is a table summarizing your discoveries: 1 / 1점

| | |
|---|---|
| Overall dev set error | 15.3% |
| Errors due to incorrectly labeled data | 4.1% |
| Errors due to foggy pictures | 3.0% |
| Errors due to partially occluded elements. | 7.2% |

| | |
|---|---|
| Errors due to partially occluded elements. | 7.2% |
| Errors due to other causes | 1.0% |

In this table, 4.1%, 7.2%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about 7.2/15.3 = 47% of your errors are due to partially occluded elements.

You find out that there is an anti-reflective film guarantee to eliminate the sun reflection, but it is quite costly. Which of the following gives the best description of what the investment in the film can do to the model?

- The overall test set error will be reduced by at most 7.2%.
- The film will reduce at least 7.2% of the dev set error.
- The film will reduce the dev set error with 7.2% at the most.

더 보기

맞습니다
Yes. Remember that this 7.2% gives us an estimate for the ceiling of how much the error can be reduced when the cause is fixed.

**10.** You decide to use data augmentation to address foggy images. You find 1,000 pictures of fog off the internet, and "add" them to clean images to synthesize foggy days, like this: 1 / 1점


We can't use this data since they have a different distribution from the ones we used (internet and front-facing camera). True/False?

- False
- True

더 보기

맞습니다
Correct. The new synthesized images are added to the training set and as long as they look realistic to the human eye this will be useful data to train the model.

**11.** After working further on the problem, you've decided to correct the incorrectly labeled data. Your team corrects the labels of the wrongly predicted images on the dev set. 1 / 1점

You have to correct the labels of the test so test and dev sets have the same distribution, but you won't change the labels on the train set because most models are robust enough they don't get severely affected by the difference in distributions. True/False?

- True, as pointed out. we must keep dev and test with the same distribution. And the labels at training should be fixed only in case of a systematic error.
- False, the test set should be changed. but also the train set to keep the same distribution between the train. dev. and test sets.
- False, the test set shouldn't be changed since we want to know how the model performs in real data.

더 보기

맞습니다
Correct! To successfully train a model, the dev set and test set should come from the same distribution. Also, the deep learning models are robust enough to handle a small change in distributions, but if the errors are systematic they can significantly affect the training of the model.

**12.** One of your colleagues at the startup is starting a project to classify road signs as stop, dangerous curve, construction ahead, dead-end, and speed limit signs. Given how specific the signs are, he has only a small dataset and hasn't been able to create a good model. You offer your help providing the trained weights (parameters) of your model to transfer knowledge. 1 / 1점

But your colleague points out that his problem has more specific items than the ones you used to train your model. This makes the transfer of knowledge impossible. True/False?

- True
- False

더 보기

맞습니다
Correct. The model can benefit from the pre-trained model since there are many features learned by your model that can be used in the new problem.

**13.** One of your colleagues at the startup is starting a project to classify stop signs in the road as speed limit signs or not. He has approximately 30,000 examples of each image and 30,000 images without a sign. He thought of using your model and applying transfer learning but then he noticed that you use multi-task learning, hence he can't use your model. True/False? 1 / 1점

- False
- True

더 보기

맞습니다
Correct. When using transfer learning we can remove the last layer. That is one of the aspects that is different from a binary classification problem.

**14.** To recognize red and green lights, you have been using this approach: 1 / 1점

- **(A)** Input an image (x) to a neural network and have it directly learn a mapping to make a prediction as to whether there's a red light and/or green light (y).

A teammate proposes a different, two-step approach:

- **(B)** In this two-step approach, you would first (i) detect the traffic light in the image (if any), then (ii) determine the color of the illuminated lamp in the traffic light.

Between these two, Approach B is more of an end-to-end approach because it has distinct steps for the input end and the output end. True/False?

- True
- False

더 보기

맞습니다
Yes. (A) is an end-to-end approach as it maps directly the input (x) to the output (y).

**15.** Consider the following two approaches, A and B: 1 / 1점

- **(A)** Input an image (x) to a neural network and have it directly learn a mapping to make a prediction as to whether there's a red light and/or green light (y).
- **(B)** In this two-step approach, you would first (i) detect the traffic light in the image (if any), then (ii) determine the color of the illuminated lamp in the traffic light.

Approach A tends to be more promising than approach B if you have a _____ (fill in the blank).

- Large training set
- Problem with a high Bayes error.
- Multi-task learning problem.
- Large bias problem.

더 보기

맞습니다
Yes. In many fields, it has been observed that end-to-end learning works better in practice, but requires a large amount of data.