

## Try again once you are ready

Grade  
received **66.66%**

Latest Submission  
Grade 66.67%

To pass 80% or  
higher

Try again

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 a task of what leading a machine learning project could be like!

1 / 1 point

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, the above 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, that could be helpful for training even if the distribution of internet data is not the same.

You are just getting started on 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).

- ☒ Spend a few days training a basic model and see what mistakes it makes.
- ☐ 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.
- ☐ Spend a few days getting the internet data, so that you understand better what data is available.
- ☐ Spend a few days checking what is human-level performance for these tasks so that you can get an accurate estimate of Bayes error.

Expand

Correct

As discussed in lecture, applied ML is a highly iterative process. If you train a basic model and carry out error analysis (see what mistakes it makes) it will help point you in more promising directions.

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.

0 / 1 point

Suppose that you use a sigmoid function for the output layer, and the output  $\hat{y}$  has shape (5, 1). Which of the following best describes the cost function?

- ☐  $\frac{\exp \hat{y}_j^{(i)}}{\sum_{j=1}^5 \exp \hat{y}_j^{(i)}}$
- ☐  $\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^5 \mathcal{L}(\hat{y}_j^{(i)}, y_j^{(i)})$
- ☐  $\frac{1}{m} \sum_{i=1}^m \left( -y_j^{(i)} \log \hat{y}_j^{(i)} - (1 - y_j^{(i)}) \log (1 - \hat{y}_j^{(i)}) \right)$

Expand

Incorrect

You did not choose an option.

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 / 1 point

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

[Expand](#)

**Correct**

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 / 1 point

- 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?

- ☒ False
- ☐ True

[Expand](#)

**Correct**

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 different distribution than the images you were able to find and download off the internet. Which of the following are true about the train/dev/test split?

0 / 1 point

- ☐ The dev and test set must come from the front-facing camera.
- ☒ The dev and test sets must come from the same distribution.

**Correct**

Correct. This is required to aim the target where we want to be.

- ☐ The dev and test sets must contain some images from the internet.
- ☒ The train, dev, and test must come from the same distribution.

**This should not be selected**

Most Deep Learning models are robust enough to allow for slightly different distributions between train and dev/test.

[Expand](#)



**Incorrect**

You didn't select all the correct answers

6. Assume you've finally chosen the following split between the data:

0 / 1 point

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



Expand



**Incorrect**

The training-dev error and the dev error are not that different to come to this conclusion.

7. Assume you've finally chosen the following split between the data:

0 / 1 point

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)	2%
Training-Dev	20,000 images randomly picked from (900,000 internet images + 60,000 car's front-facing camera images)	2.3%
Dev	20,000 images from your car's front-facing camera	1.3%
Test	20,000 images from the car's front-facing camera	1.1%

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 harder than the dev/test distribution. What do you think?

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



Expand



**Incorrect**

Notice that the test and dev errors are lower than the train and train-dev errors.

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

1 / 1 point

Overall dev set error	15.3%
-----------------------	-------

Errors due to incorrectly labeled data	4.1%
Errors due to foggy pictures	8.0%
Errors due to rain drops stuck on your car's front-facing camera	2.2%
Errors due to other causes	1.0%

In this table, 4.1%, 8.0%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about  $8.0/15.3 = 52\%$  of your errors are due to foggy pictures.

The results from this analysis implies that the team's highest priority should be to bring more foggy pictures into the training set so as to address the 8.0% of errors in that category. True/False?

Additional note: there are subtle concepts to consider with this question, and you may find arguments for why some answers are also correct or incorrect. We recommend that you spend time reading the feedback for this quiz, to understand what issues that you will want to consider when you are building your own machine learning project.

- ☐ True because it is greater than the other error categories added together  $8.0 > 4.1 + 2.2 + 1.0$ .
- ☒ False because it depends on how easy it is to add foggy data. If foggy data is very hard and costly to collect, it might not be worth the team's effort.
- ☐ True because it is the largest category of errors. We should always prioritize the largest category of errors as this will make the best use of the team's time.
- ☐ First start with the sources of error that are least costly to fix.

[Expand](#)

☒ **Correct**

Correct. This is the correct answer. You should consider the tradeoff between the data accessibility and potential improvement of your model trained on this additional data.

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 / 1 point

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 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 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.
- ☐ The overall test set error will be reduced by at most 7.2%.

[Expand](#)

☒ **Correct**

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 / 1 point



Which of the following statements do you agree with?

- ☐ Adding synthesized images that look like real foggy pictures taken from the front-facing camera of your car to the training dataset won't help the model improve because it will introduce avoidable bias.
- ☐ There is little risk of overfitting to the 1,000 pictures of fog so long as you are combining it with a much larger ( $>>1,000$ ) set of clean/non-foggy images.
- ☒ So long as the synthesized fog looks realistic to the human eye, you can be confident that the synthesized data is accurately capturing the distribution of real foggy images (or a subset of it), since human vision is very accurate for the problem you're solving.

 Expand

 **Correct**

Yes. If the synthesized images look realistic, then the model will just see them as if you had added useful data to identify road signs and traffic signals in foggy weather. I will very likely help.

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. Which of the following is a necessary step to take?

1 / 1 point

- ☐ Create a train-dev set to estimate how many incorrectly labeled examples are in the train set.
- ☐ Correct the labels of the train set.
- ☐ Use a correctly labeled version and an incorrectly labeled version to make the model more robust.
- ☒ Correct the labels of the test set.

 Expand

 **Correct**

Correct. Recall that the dev set and the test set must come from the same distribution.

12. So far your algorithm only recognizes red and green traffic lights. One of your colleagues in the startup is starting to work on recognizing a yellow traffic light. (Some countries call it an orange light rather than a yellow light; we'll use the US convention of calling it yellow.) Images containing yellow lights are quite rare, and she doesn't have enough data to build a good model. She hopes you can help her out using transfer learning.

1 / 1 point

What do you tell your colleague?

- ☐ You cannot help her because the distribution of data you have is different from hers, and is also lacking the yellow label.
- ☐ If she has (say) 10,000 images of yellow lights, randomly sample 10,000 images from your dataset and put your and her data together. This prevents your dataset from "swamping" the yellow lights dataset.
- ☐ Recommend that she try multi-task learning instead of transfer learning using all the data.
- ☒ She should try using weights pre-trained on your dataset, and fine-tuning further with the yellow-light dataset.

 Expand

 **Correct**

Yes. You have trained your model on a huge dataset, and she has a small dataset. Although your labels are different, the parameters of your model have been trained to recognize many characteristics of road and traffic images which will be useful for her problem. This is a perfect case for transfer learning, she can start with a model with the same architecture as yours, change what is after the last hidden layer and initialize it with your trained parameters.

13. Another colleague wants to use microphones placed outside the car to better hear if there are other vehicles around you. For example, if there is a police vehicle behind you, you would be able to hear their siren. However, they don't have much to train this audio system. How can you help?

1 / 1 point

- ☒ Neither transfer learning nor multi-task learning seems promising.
- ☐ Multi-task learning from your vision dataset could help your colleague get going faster. Transfer learning seems significantly less promising.
- ☐ Either transfer learning or multi-task learning could help our colleague get going faster.
- ☐ Transfer learning from your vision dataset could help your colleague get going faster. Multi-task learning seems significantly less promising.

 Expand

 **Correct**

Yes. The problem he is trying to solve is quite different from yours. The different dataset structures make it probably impossible to use transfer learning or multi-task learning.

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

1 / 1 point

- **(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?

- ☒ False
- ☐ True

 Expand

 **Correct**

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

15. To recognize a stop sign you use the following approach:

0 / 1 point

First, we localize any traffic sign in an image. After that, we determine if the sign is a stop sign or not.

This is a better approach than an end-to-end model for which of the following cases? Choose the best answer.

- ☒ There is a large amount of data.
- ☐ There is not enough data to train a big neural network.
- ☐ The problem has a high Bayes error.
- ☐ There are available models which we can use to transfer knowledge.

 Expand

 **Incorrect**

No. In this case, it is more promising to use an end-to-end approach.