

Congratulations! You passed!

[Go to next item](#)

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



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

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 checking what is human-level performance for these tasks so that you can get an accurate estimate of Bayes error.
- Spend a few days getting the internet data, so that you understand better what data is available.
- 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.

[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. For the output layer, a softmax activation would be a good choice for the output layer because this is a multi-task learning problem. True/False? 1 / 1 point

- False
- True

[Expand](#)

Correct

Softmax would be a good choice if one and only one of the possibilities (stop sign, speed bump, pedestrian crossing, green light and red light) was present in each image.

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

1 / 1 point

- 500 images of the test set, on which the algorithm made a mistake.
- 500 images of the training-dev set, on which the algorithm made a mistake.
- 500 images of the train set, on which the algorithm made a mistake.
- 500 images of the dev set, on which the algorithm made a mistake.

 Expand

 Correct

Correct. We focus on images that the algorithm got wrong from the dev set. That is the one we use to make choices between different iterations of the system.

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.

When using a non fully labeled image such as $y^{(i)} = \begin{bmatrix} 0 \\ ? \\ 1 \\ ? \\ 1 \end{bmatrix}$, which of the following strategies is most appropriate to calculate the loss function to train as a multi-task learning problem?

- Make the missing entries equal to 0.
- Make the missing entries equal to 1.
- Calculate the loss as $\sum_{j=1}^L \hat{y}_j y_j$ where the sum goes over all the known components of $y^{(i)}$.
- It is not possible to use non fully labeled images if we train as a multi-task learning problem.

 Expand

 Correct

Correct. We can't use the components of the labels that are missing but we can use the ones we have to train the model.

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. The best way to split the data is using the 900,000 internet images to train, and divide the 100,000 images from your car's front-facing camera between dev and test sets. True/False?

1 / 1 point

- True
- False

 Expand

 Correct

Correct. 100,000 images are too many to use in dev and test. A better distribution would be to use 80,000 of those images to train, and split the rest between dev and test.

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

1 / 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)	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%. Which of the following are True? (Check all that apply).

You have a large avoidable-bias problem because your training error is quite a bit higher than the human-level error.

 Correct

You have a large variance problem because your training error is quite higher than the human-level error.

Your algorithm overfits the dev set because the error of the dev and test sets are very close.

You have a large data-mismatch problem because your model does a lot better on the training-dev set than on the dev set

 Correct

You have a large variance problem because your model is not generalizing well to data from the same training distribution but that it has never seen before.

 Expand

 Correct

Great, you got all the right answers.

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

1 / 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)	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?

There's insufficient information to tell if your friend is right or wrong.

Your friend is wrong. (I.e., Bayes error for the training data distribution is probably higher than for the dev/test distribution.)

Your friend is right. (I.e., Bayes error for the training data distribution is probably lower than for the dev/test distribution.)

 Expand

 Correct

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 / 1 point

Overall dev set error	15.3%
Errors due to incorrectly labeled data	4.1%
Errors due to foggy pictures	2.0%
Errors due to partially occluded elements.	8.2%
Errors due to other causes	1.0%

In this table, 4.1%, 8.2%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about $8.2/15.3 = 54\%$ of your errors are due to partially occluded elements in the image.

Which of the following is the correct analysis to determine what to prioritize next?

- Since $8.2 > 4.1 + 2.0 + 1.0$, the priority should be to get more images with partially occluded elements.
- You should weigh how costly it would be to get more images with partially occluded elements, to decide if the team should work on it or not.
- You should prioritize getting more foggy pictures since that will be easier to solve.
- Since there is a high number of incorrectly labeled data in the dev set, you should prioritize fixing the labels on the whole training set.

 Expand

 Correct

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

9. You can buy a specially designed windshield wiper that helps wipe off some of the raindrops on the front-facing camera.

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%

Which of the following statements do you agree with?

- 2.2% would be a reasonable estimate of how much this windshield wiper will improve performance.
- 2.2% would be a reasonable estimate of the minimum amount this windshield wiper could improve performance.
- 2.2% would be a reasonable estimate of the maximum amount this windshield wiper could improve performance.
- 2.2% would be a reasonable estimate of how much this windshield wiper could worsen performance in the worst case.

 Expand

 Correct

Yes. You will probably not improve performance by more than 2.2% by solving the raindrops problem. If your dataset was infinitely big, 2.2% would be a perfect estimate of the improvement you can achieve by purchasing a specially designed windshield wiper that removes the raindrops.

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



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

True

False

 Expand

 Correct

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

1 / 1 point

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

 Expand

 Correct

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

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 / 1 point

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

 Expand

 Correct

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 road signs as stop, dangerous curve, construction ahead, dead-end, and speed limit signs. He has approximately 30,000 examples of each image and 30,000 images without a sign. This case could benefit from using multi-task learning. True/False?

1 / 1 point

False

True

 Expand

 Correct

Correct. There are a lot of high-level features that all the required signs share. This is a great scenario to make use of 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?

True

False

 Expand

 Correct

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 / 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).
- (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 bias problem.

Large training set

Multi-task learning problem.

Problem with a high Bayes error.

 Expand

Correct

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