Your grade: 80%

Your latest: **80%** • Your highest: **80%** • To pass you need at least 80%. We keep your highest score.

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 an idea 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, 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).

- (●) Train a basic model and do error analysis.
- ( ) Spend some time searching the internet for the data most similar to the conditions you expect on production.
- ( ) Invest a few days in thinking on potential difficulties, and then some more days brainstorming about possible solutions, before training any model.
- ( ) 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.

Correct
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. 1 / 1 point

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?

- (●) False
- ( ) True

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

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.

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

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. Which of the following are true about the train/dev/test split? 1 / 1 point

- [ ] The train, dev, and test must come from the same distribution.
- [x] 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.

- [x] The dev and test set must come from the front-facing camera.

Correct
Correct. This is the distribution we care about most, thus we should use this as a target.

- [ ] The dev and test sets must contain some images from the internet.

6. Assume you've finally chosen the following split between the data: 0 / 1 point

- ○ You have a high variance problem.
- ○ You have a high bias.
- ○

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?

- ○ The size of the train-dev set is too high.
- ◉ You have a large data-mismatch problem.

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

- ◉ 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.)
- ○

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.

Incorrect
The higher error on the dev/test set compared to the training set suggests that the model performs worse on the dev/test distribution, but without additional information, such as human-level error on both distributions, it's unclear if this is due to the distribution's difficulty or other factors like overfitting.

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.

- ○ First start with the sources of error that are least costly to fix.
- ○ True because it is the largest category of errors. We should always prioritize the largest category of

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.

- ○ 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.

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 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 the minimum amount this windshield wiper could improve performance.
- ○ 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 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.

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

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. **0 / 1 point**

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?

- ◉ False, the test set shouldn't be changed since we want to know how the model performs in real data.
- ○ False, the test set should be changed, but also the train set to keep the same distribution between the train, dev, and test sets.
- ○ 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.

Incorrect
When correcting the labels we are not creating false data we are eliminating unreal data from the set.

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

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 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 **1 / 1 point**

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?

- (●) False
- ( ) True

Correct
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 / 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

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

15. An end-to-end approach doesn't require that we hand-design useful features, it only requires a large enough model. True/False? **1 / 1 point**

- (●) True
- ( ) False

Correct
Correct. This is one of the major characteristics of deep learning models, that we don't need to hand-design the features.