Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

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 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 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.
- ( ) 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.

Expand

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. For the output layer, which of the following gives you the most appropriate activation function? 1 / 1 point

- ( ) ReLU
- ( ) Softmax
- (●) Sigmoid
- ( ) Linear

Expand

Correct
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. When trying to determine what strategy to implement to improve the performance of a model, we manually check all images of the training set where the algorithm was successful. True/False? 1 / 1 point

- True
- False

Expand

Correct
Correct. This set should be too large to manually check all the images. It is better to focus on the images that the algorithm got wrong from the dev set. Also, choose a large enough subset that we can manually check.

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, when an image is not fully labeled (for example: $\begin{pmatrix} 0 \\ ? \\ ? \\ 1 \\ 0 \end{pmatrix}$) we can use it if we ignore those entries when calculating the loss function. True/False?

- True
- False

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

- [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.

☑ 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.

☐ The train, dev, and test must come from the same distribution.

Expand

✓ **Correct**
Great, you got all the right answers.

6. 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)	12%
Training-Dev	20,000 images randomly picked from (900,000 internet images + 60,000 car's front-facing camera images)	15.1%
Dev	20,000 images from your car's front-facing camera	12.6%
Test	20,000 images from the car's front-facing camera	15.8%

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 variance problem.

○ You have a large data-mismatch problem.

◉ You have a high bias.

○ You have a too low avoidable bias.

Expand

✓ **Correct**
Correct. The avoidable bias is significantly high since the training error is a lot higher than the human-level error.

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)	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 you conclude that the Bayes error for the dev/test distribution is probably higher than for the train distribution. True/False?

- True
- False

Expand

Correct

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.

- 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 greater than the other error categories added together $8.0 > 4.1 + 2.2 + 1.0$.
- 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 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 could worsen performance in the worst case.
- 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 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.

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

Which of the following do you agree with?

- If used, the synthetic data should be added to the training/dev/test sets in equal proportions.
- If used, the synthetic data should be added to the training set.
- With this technique, we duplicate the size of the training set by synthesizing a new foggy image for each image in the training set.
- It is irrelevant how the resulting foggy images are perceived by the human eye, the most important thing is that they are correctly synthesized.

Expand

Correct
Yes. The synthetic data can help to train the model to get better performance at the dev set, but shouldn't be added to the dev or test sets because they don't represent our target in a completely accurate way.

11. After working further on the problem, you've decided to correct the incorrectly labeled data on the dev set. Which of these statements do you agree with? (Check all that apply).

1 / 1 point

- [x] You do not necessarily need to fix the incorrectly labeled data in the training set, because it's okay for the training set distribution to differ from the dev and test sets. Note that it is important that the dev set and test set have the same distribution.

Correct
True, deep learning algorithms are quite robust to having slightly different train and dev distributions.

- [ ] You should not correct the incorrectly labeled data in the test set, so that the dev and test sets continue to come from the same distribution.
- [x] You should also correct the incorrectly labeled data in the test set, so that the dev and test sets continue to come from the same distribution.

Correct
Yes because you want to make sure that your dev and test data come from the same distribution for your algorithm to make your team's iterative development process efficient.

- [ ] You should correct incorrectly labeled data in the training set as well so as to avoid your training set now being even more different from your dev set.

Expand

Correct
Great, you got all the right answers.

12. Your client asks you to add the capability to detect dogs that may be crossing the road to the system. He can provide a relatively small set containing dogs. Which of the following do you agree most with? **1 / 1 point**

- (•) You can use weights pre-trained on the original data, and fine-tune with the data now including the dogs.
- ( ) You should train a single new model for the dogs' task, and leave the previous model as it is.
- ( ) You will have to re-train the whole model now including the dogs' data.
- ( ) Using pre-trained weights can severely hinder the ability of the model to detect dogs since they have too many learned features.

Expand

Correct
Correct. Since your model has learned useful low-level features to tackle the new task we can conserve those by using the pre-trained weights.

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 a stop sign you use the following approach:** First, you use a neural network to predict bounding box co-ordinates around all traffic signs (if any) within an input image. You then pass the results to a different neural network to determine if the predicted traffic signs (if any) are a stop sign or not. We are using multi-task learning. True/False? **1 / 1 point**

- ( ) True

- (●) False

Expand

Correct
Correct. Multi-task learning is about joining several tasks that can benefit from each other. Since there are 2 different neural networks being used here that do not share weights (i.e. structure), this problem has 2 single task learning neural networks and not a multi-task learning setup.

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**

- ( ) False
- (●) True

Expand

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