Congratulations! You passed!

Grade received 93.33% **Latest Submission Grade** 93.33% **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.

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

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, 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).

- ( ) 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.
- (●) Train a basic model and do error analysis.

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

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. 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

- 500 randomly chosen images
- 10,000 randomly chosen images
- 10,000 images on which the algorithm made a mistake
- 500 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. How should you split the dataset into train/dev/test sets?

1 / 1 point

- 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 80,000 images from your car's front-facing camera. The 20,000 remaining images will be split equally in dev and test sets.

- [ ] 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.

Expand

Correct
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 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 variance problem because your model is not generalizing well to data from the same training distribution but that it has never seen before.
- [x] 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.
- [x] 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

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)	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.

- ○ 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.
- ○ 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 errors as this will make the best use of the team's time.

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 the dev set error with 7.2% at the most.
- ○ 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.

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 do you agree with?

- ( ) 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.
- (•) If used, the synthetic data should be added to the training set.
- ( ) If used, the synthetic data should be added to the training/dev/test sets in equal proportions.

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. Your team corrects the labels of the wrongly predicted images on the dev set.

1 / 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.

Expand

Correct
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. 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

- ( ) Using pre-trained weights can severely hinder the ability of the model to detect dogs since they have too many learned features.
- ( ) You will have to re-train the whole model now including the dogs' data.

- You should train a single new model for the dogs' task, and leave the previous model as it is.
- (selected) You can use weights pre-trained on the original data, and fine-tune with the data now including the dogs.

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 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?

0 / 1 point

- (selected) True
- False

Expand

Incorrect
When using transfer learning we can remove the last layer, which 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?

- (selected) False
- True

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).

- (selected) Large training set
- Problem with a high Bayes error.

- Large bias problem.
- Multi-task learning problem.

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.