Congratulations! You passed!

Grade received 80% **Latest Submission Grade** 80% **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 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

$$y^{(i)} = \begin{array}{cl} \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{array}\right] & \begin{array}{l} \text{“stop sign”} \\ \text{“pedestrian crossing sign”} \\ \text{“construction ahead sign”} \\ \text{“red traffic light”} \\ \text{“green traffic light”} \end{array} \end{array}$$

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 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 training a basic model and see what mistakes it makes.
- 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. 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

- True
- False

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?

0 / 1 point

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

Expand

Incorrect
We should avoid this since we might overfit the test set.

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?

- True
- (selected) False

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

- 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.
- 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.
- (selected) 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.

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

Expand

Incorrect
The train-dev set has an adequate size to estimate the true error in the training set.

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 $8.0 > 4.1 + 2.2 + 1.0$.
- ○ 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.
- ◉ 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.

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%

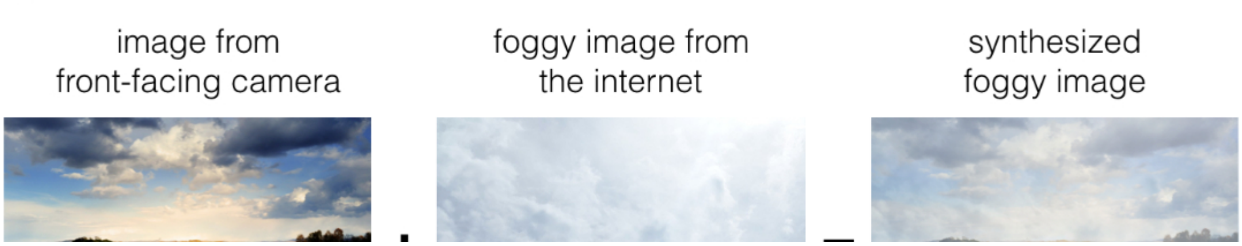
Which of the following statements do you agree with?

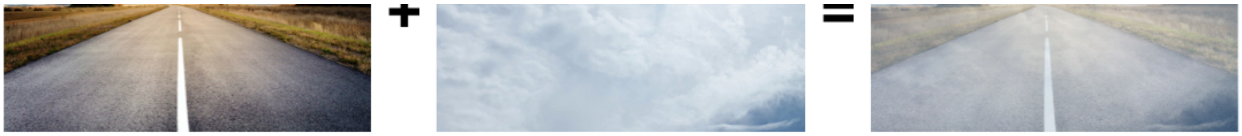
- ◉ 2.2% would be a reasonable estimate of the maximum amount this windshield wiper could improve performance.
- ○ 2.2% would be a reasonable estimate of how much this windshield wiper could worsen performance in the worst case.
- ○ 2.2% would be a reasonable estimate of how much this windshield wiper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

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: **0 / 1 point**





Which of the following do you agree with?

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

Expand

Incorrect
No. Our objective is to have images that look realistic to the human eye.

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 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.
- False, the test set shouldn't be changed since we want to know how the model performs in real data.

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. 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?

- She should try using weights pre-trained on your dataset, and fine-tuning further with the yellow-light dataset.
- Recommend that she try multi-task learning instead of transfer learning using all the data.
- 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.

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

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

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. When building a system to detect cattle crossing a road from images taken with the front-facing camera of a truck, the designers had a large dataset of images. Which of the following might be a reason to use an end-to-end approach?

1 / 1 point

- ○ It requires less computational resources.
- ○ That is the default approach on computer vision tasks.
- ○ This approach will make use of useful hand-designed components.
- ◉ There is a large dataset available.

Expand

Correct

Correct. To get good results when using an end-to-end approach, it is necessary to have a big dataset.

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

Expand

Correct

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