DLS/C3 / W2/ QUIZ

Autonomous Driving (Case Study)

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!

   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

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

   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, 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 getting the internet data, so that you understand better what data is available.
- ○ 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 checking what is human-level performance for these tasks so that you can get an accurate estimate of Bayes error.
- ◉ Spend a few days training a basic model and see what mistakes it makes.

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, which of the following gives you the most appropriate activation function?

- ○ ReLU
- ◉ Sigmoid
- ○ Softmax
- ○ 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?

- ○ 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:

- 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
- ◉ False

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?

- ○ 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.
- ◉ 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:

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 bias.
- ( ) You have a too low avoidable bias.
- ( ) You have a large data-mismatch problem.
- ( ) You have a high variance problem.

✓ **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:

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, a friend thinks that the training data distribution is much harder than the dev/test distribution. What do you think?

- ( ) Your friend is wrong. (i.e., Bayes error for the dev/test distribution is probably higher than for the train distribution.)
- ( ) Your friend is probably right. (i.e., Bayes error for the dev/test distribution is probably lower than for the train distribution.)
- (●) There's insufficient information to tell if your friend is right or wrong.

✗ **Incorrect**
Notice that the test and dev errors are lower than the train and train-dev errors.

correct answer: B

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:

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.
- 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 decide to focus on the dev set and check by hand what the errors are due to. Here is a table summarizing your discoveries:

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 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.
- ◉ The film will reduce the dev set error with 7.2% at the most.

✓ **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:

Which of the following do you agree with?

- ◉ If used, the synthetic data should be added to 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.
- ○ With this technique, we duplicate the size of the training set by synthesizing a new foggy image for each image in the training set.

✓ **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).

- [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.

- [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.

✓ **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?

- ( ) 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.
- ( ) Using pre-trained weights can severely hinder the ability of the model to detect dogs since they have too many learned features.
- (•) 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?

- ○ True
- ◉ False

Expand

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:

- **(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?

- ◉ False
- ○ True

Expand

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?

- ○ True
- ◉ False

Expand

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

Correct answer: True. Because end-to-end has 2 major points: large model + do not require hand-design features.