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 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.
- ( ) 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.

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, a softmax activation would be a good choice for the output layer because this is a multi-task learning problem. True/False? **0 / 1 point**

- (•) True
- ( ) False

Expand

Incorrect

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**

- (•) False

- True

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.

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 0.
- Make the missing entries equal to 1.
- 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.

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**

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

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 too low avoidable bias.
- ◉ You have a high bias.
- ○ You have a large data-mismatch problem.

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 the errors are due to. Here is a table summarizing your discoveries: 0 / 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 shouldn't invest all your efforts to get more images with partially occluded elements since 4.1 + 3.0 + 1.0 = 8.1

> 7.2. True/False?

- (●) True
- ( ) False

Expand

✗ **Incorrect**
Incorrect. These kinds of arguments don't help us to decide on the strategy to follow. Other factors should be used, such as the tradeoff between the cost of getting new images and the improvement of the system performance.

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 at least 7.2% of the dev set error.
- (●) 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%.

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.
- ( ) 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.
- ( ) 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). **0 / 1 point**

- [ ] 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 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.

  ! **This should not be selected**
  No, deep learning algorithms are quite robust to having slightly different train and dev distributions.
- [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 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.

Expand

Incorrect
You didn't select all the correct answers

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

Expand

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 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. 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

- 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.
- It requires less computational resources.

Expand

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

15. To recognize a stop sign you use the following approach: 1 / 1 point

First, we localize any traffic sign in an image. After that, we determine if the sign is a stop sign or not.

This is a better approach than an end-to-end model for which of the following cases? Choose the best answer.

- There is not enough data to train a big neural network.
- There are available models which we can use to transfer knowledge.
- There is a large amount of data.
- The problem has a high Bayes error.

Expand

Correct
Correct. This might be the most important factor when deciding whether to use an end-to-end approach.