Try again once you are ready

Grade received 66.66% | Latest Submission Grade 53.33% | To pass 80% or higher

Try again

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!

0 / 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).

- (selected) Spend some time searching the internet for the data most similar to the conditions you expect on production.
- Invest a few days in thinking on potential difficulties, and then some more days brainstorming about possible solutions, before training any model.
- 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.

Expand

Incorrect
Doing a detailed error analysis of a basic model can help us determine confidently if the system will truly benefit from the extra data.

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
- (selected) 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

- 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.
- (selected) 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
The objective of error analysis is to improve the performance on the dev/test set.

4. After working on the data for several weeks, your team ends up with the following data: **0 / 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?

- (selected) False
- True

Expand

Incorrect
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. 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.
- 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 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:

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

Expand

Correct
Correct. Since the difference between the training-dev error and the training error is high.

7. 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)	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%. Based on the information given, a friend thinks that the training data distribution is much easier than the dev/test distribution. What do you think?

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

Expand

Incorrect

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	2.0%

Errors due to partially occluded elements.	8.2%
Errors due to other causes	1.0%

In this table, 4.1%, 8.2%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about 8.2/15.3 = 54% of your errors are due to partially occluded elements in the image.

Which of the following is the correct analysis to determine what to prioritize next?

- ○ You should weigh how costly it would be to get more images with partially occluded elements, to decide if the team should work on it or not.
- ○ You should prioritize getting more foggy pictures since that will be easier to solve.
- ◉ Since there is a high number of incorrectly labeled data in the dev set, you should prioritize fixing the labels on the whole training set.
- ○ Since 8.2 > 4.1 + 2.0 + 1.0, the priority should be to get more images with partially occluded elements.

Expand

Incorrect
This is most likely going to be too costly and with only marginal benefits for the project.

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:

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 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?

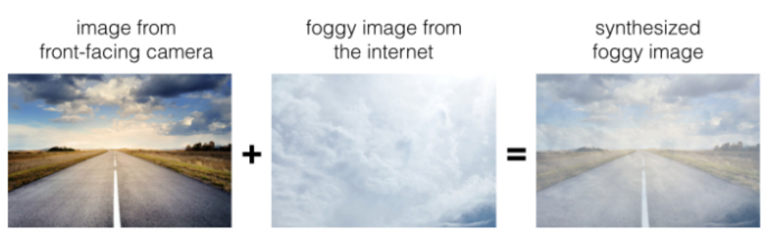
- ◉ The overall test set error will be reduced by at most 7.2%.
- ○ The film will reduce at least 7.2% of the dev set error.
- ○ The film will reduce the dev set error with 7.2% at the most.

Expand

Incorrect
We can't extrapolate the dev set error to the test error.

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



We can't use this data since they have a different distribution from the ones we used (internet and front-facing camera). True/False?

- ◉ False
- ○ True

Expand

Correct
Correct. The new synthesized images are added to the training set and as long as they look realistic to the human eye this will be useful data to train the model.

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

- [x] 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.

! **This should not be selected**
No, because you have to correct the incorrectly labeled data in the test set to make sure that your test set and dev set 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.
- [ ] 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

- (•) True
- ( ) False

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
- [x] 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

- [x] 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.