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

- ◉ False
- ○ True

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 training-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 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:

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?

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

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. 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 80,000 images from your car's front-facing camera. The 20,000 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 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).

- [ ] Your algorithm overfits the dev set because the error of the dev and test sets are very close.
- [ ] You have a large variance problem because your training error is quite higher than the human-level error.
- [ ] 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

- [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)	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.)

- [x] There's insufficient information to tell if your friend is right or wrong.

- [ ] Your friend is right. (I.e., Bayes error for the training data distribution is probably lower than for the dev/test distribution.)

Expand

Correct
The algorithm does better on the distribution of data it trained on. But you don't know if it's because it trained on that distribution or if it really is easier. To get a better sense, measure human-level error separately on both distributions.

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 $8.0 > 4.1 + 2.2 + 1.0$.
- [ ] 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.
- [x] 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 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 in the image.

From this table, we can conclude that if we fix the incorrectly labeled data we will reduce the overall dev set error to 11.2%. True/False?

- [x] False

- [ ] True

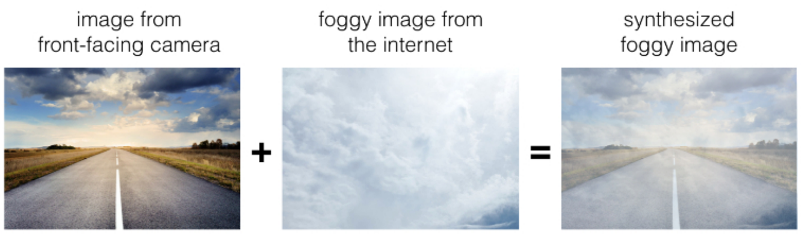
Expand

Correct

Correct. The 4.1 only gives you an estimate of the ceiling of how much the error can be improved by fixing the labels.

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 statements do you agree with?

- There is little risk of overfitting to the 1,000 pictures of fog so long as you are combining it with a much larger (>>1,000) set of clean/non-foggy images.
- Adding synthesized images that look like real foggy pictures taken from the front-facing camera of your car to the training dataset won't help the model improve because it will introduce avoidable bias.
- So long as the synthesized fog looks realistic to the human eye, you can be confident that the synthesized data is accurately capturing the distribution of real foggy images (or a subset of it), since human vision is very accurate for the problem you're solving.

Expand

Correct

Yes. If the synthesized images look realistic, then the model will just see them as if you had added useful data to identify road signs and traffic signals in foggy weather. I will very likely help.

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. Which of the following is a necessary step to take?

1 / 1 point

- Correct the labels of the test set.
- Correct the labels of the train set.
- Create a train-dev set to estimate how many incorrectly labeled examples are in the train set.
- Use a correctly labeled version and an incorrectly labeled version to make the model more robust.

Expand

Correct

Correct. Recall that the dev set and the test set must come from the same distribution.

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?

model. This makes the transfer of knowledge impossible. True/False?

- False
- True

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?

0 / 1 point

- False
- True

Expand

Incorrect
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, we localize any traffic sign in an image. After that, we determine if the sign is a stop sign or not. We are using multi-task learning. True/False?

1 / 1 point

- True
- False

Expand

Correct
Correct. Multi-task learning is about joining several tasks that can benefit from each other.

15. An end-to-end approach doesn't require that we hand-design useful features, it only requires a large enough model. True/False?

0 / 1 point

- False
- True

Expand

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