Congratulations! You passed!

Grade received 86.66% | Latest Submission Grade 86.67% | 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

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

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 checking what is human-level performance for these tasks so that you can get an accurate estimate of Bayes error.
- ○ 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.

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. 1 / 1 point

Suppose that you use a sigmoid function for the output layer, and the output $\hat{y}$ has shape (5, 1). Which of the following best describes the cost function?

- ○ $\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{5}\mathcal{L}(\hat{y}_i^{(j)}, y_i^{(j)})$
- ○ $\frac{\exp \hat{y}_j^{(i)}}{\sum_{j=1}^{5}\exp \hat{y}_j^{(i)}}$
- ○ $\frac{1}{m}\sum_{i=1}^{m}\left(-y^{(i)}\log\hat{y}^{(i)} - (1-y^{(i)})\log(1-\hat{y}^{(i)})\right)$

Expand

Correct
Correct. Here we compare each component of the prediction $\hat{y}$ with the respective component of the label y, and sum over the individual losses.

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

- ○ 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: 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, 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?

- (•) True
- ( ) False

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 20,000 images from your car's front-facing camera. The 80,000 remaining images will be split equally in dev and test sets.
- (•) 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.

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)	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 large data-mismatch problem.

- ○ You have a too low avoidable bias.
- ○ You have a high variance problem.
- ◉ You have a high bias.

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

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 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 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.
- ○ 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.

Expand

✗ Incorrect
The choice should be made taking into consideration other aspects, not just this.

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.

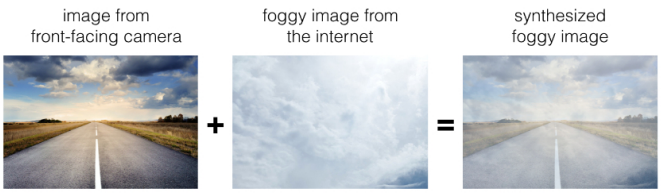
You find out that there is an anti-reflective film guarantee to eliminate the sun reflection, but it is quite costly. Which of the following gives the best description of what the investment in the film can do to the model?

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

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: 0 / 1 point



Which of the following statements do you agree with?

- ○ 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.
- ◉ 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.

Expand

✗ Incorrect

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). 1 / 1 point

- ☑ 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 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 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.

Expand

✓ Correct
Great, you got all the right answers.

Great, you got all the right answers.

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?

- ○ You cannot help her because the distribution of data you have is different from hers, and is also lacking the yellow label.
- ◉ 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.
- ○ 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. 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
- ◉ 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. 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.
- ○ The problem has a high Bayes error.
- ○ There is a large amount of data.
- ○ There are available models which we can use to transfer knowledge.

Expand

Correct

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