Congratulations! You passed!

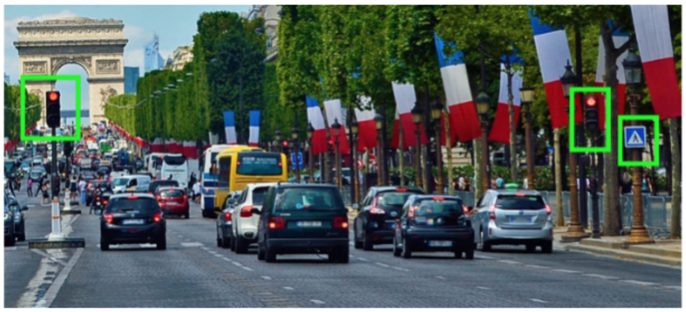
Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

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



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

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, 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 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 collecting more data using the front-facing camera of your car, to better understand how much data per unit time you can collect.

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?

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

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

- (selected) 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.

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?

- False
- (selected) True

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.
- 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.
- 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.
- (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)	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?

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

Correct
Correct. 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 can buy a specially designed windshield wiper that helps wipe off some of the raindrops on the front-facing camera. **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%

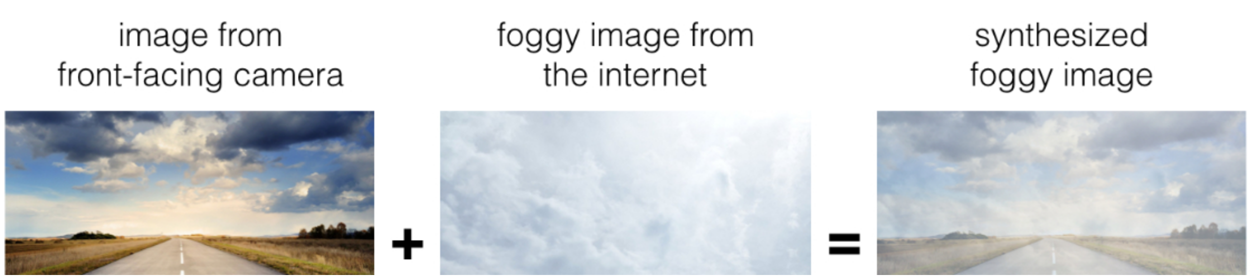
Which of the following statements do you agree with?

- 2.2% would be a reasonable estimate of the minimum amount this windshield wiper could improve performance.
- 2.2% would be a reasonable estimate of how much this windshield wiper will improve performance.
- 2.2% would be a reasonable estimate of the maximum amount this windshield wiper could improve performance.
- 2.2% would be a reasonable estimate of how much this windshield wiper could worsen performance in the worst case.

Expand

Correct
Yes. You will probably not improve performance by more than 2.2% by solving the raindrops problem. If your dataset was infinitely big, 2.2% would be a perfect estimate of the improvement you can achieve by purchasing a specially designed windshield wiper that removes the raindrops.

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?

- 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.
- 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.
- 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. **1 / 1 point**

You have to correct the labels of the test so test and dev sets have the same distribution, but you won't change the labels on the train set because most models are robust enough they don't get severely affected by the difference in distributions. True/False?

- False, the test set shouldn't be changed since we want to know how the model performs in real data.
- False, the test set should be changed, but also the train set to keep the same distribution between the train, dev, and test sets.
- True, as pointed out, we must keep dev and test with the same distribution. And the labels at training should be fixed only in case of a systematic error.

Expand

Correct
Correct! To successfully train a model, the dev set and test set should come from the same distribution. Also, the deep learning models are robust enough to handle a small change in distributions, but if the errors are systematic they can significantly affect the training of the model.

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?

- 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.
- 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.

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 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?

1 / 1 point

- (●) False
- ( ) True

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. 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

- ( ) This approach will make use of useful hand-designed components.
- ( ) It requires less computational resources.
- ( ) That is the default approach on computer vision tasks.
- (●) There is a large dataset available.

Expand

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

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

- (●) 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.