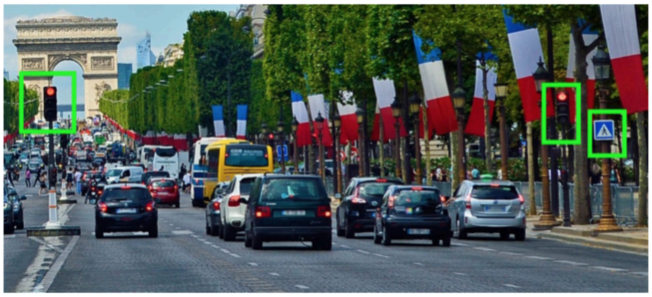
To help you practice strategies for machine learning, this week we'll present another scenario and ask how you would act. We think this "simulator" of working in a machine learning project will give a task of what leading a machine learning project could be like! 1 point

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

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 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?

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

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? 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 dev set, on which the algorithm made a mistake.
- ○ 500 images of the train set, on which the algorithm made a mistake.

4.After working on the data for several weeks, your team ends up with the following data: 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, you need to have all your $y^{(i)}$ vectors fully labeled. If one example is equal to $\begin{bmatrix} 0 \\ ? \\ 1 \\ 1 \\ ? \end{bmatrix}$ then the learning algorithm will not be able to use that example. True/False?

- ◯ False
- ◯ True

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

6.Assume you've finally chosen the following split between the data: 1 point

- ◯ You have a large data-mismatch problem.
- ◯ You have a high bias.
- ◯

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?

- ◯ The size of the train-dev set is too high.
- ◯ You have a high variance problem.

7.Assume you've finally chosen the following split between the data: 1 point

- ◯ True

- ◯ Dataset split:

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 higher than for the train distribution. True/False?

- ◯ False

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 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?

- ◯ False

- ◯ True

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 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%

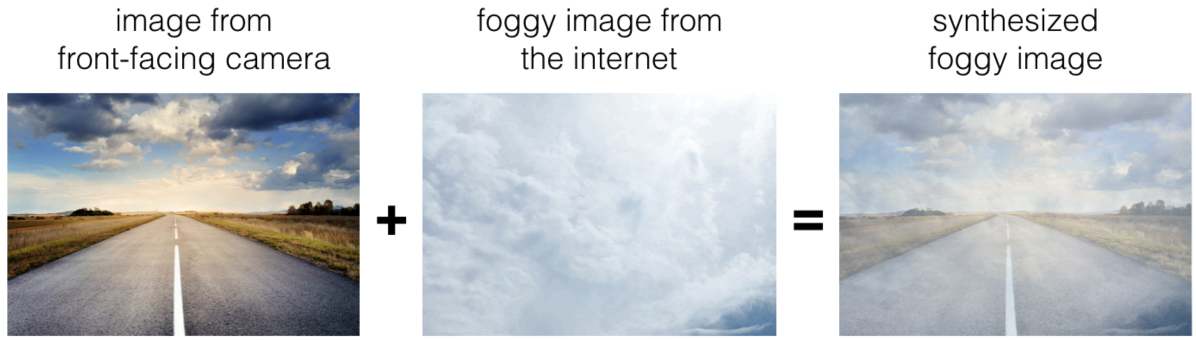
In this table, 4.1%, 7.2%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about 7.2/15.3 = 47% of your errors are due to partially occluded elements in the image.

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?

- ◯ False

- ◯ True

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 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?

- ◯ True
- ◯ False

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 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?

- ◯ 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.
- ◯ 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.

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 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.
- ◯ She should try using weights pre-trained on your dataset, and fine-tuning further with the yellow-light dataset.
- ◯ You cannot help her because the distribution of data you have is different from hers, and is also lacking the yellow label.
- ◯ Recommend that she try multi-task learning instead of transfer learning using all the data.

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 point

- ◯ False
- ◯ True

14.To recognize red and green lights, you have been using this approach: 1 point

- (A) Input an image (x) to a neural network and have it directly learn a mapping to make a prediction as to whether there's a red light and/or green light (y).

A teammate proposes a different, two-step approach:

- (B) In this two-step approach, you would first (i) detect the traffic light in the image (if any), then (ii) determine the color of the illuminated lamp in the traffic light.

  Between these two, Approach B is more of an end-to-end approach because it has distinct steps for the input end and the output end. True/False?

- ○ False
- ○ True

15. To recognize a stop sign you use the following approach: 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 is a large amount of data.
- ○ The problem has a high Bayes error.
- ○ There are available models which we can use to transfer knowledge.