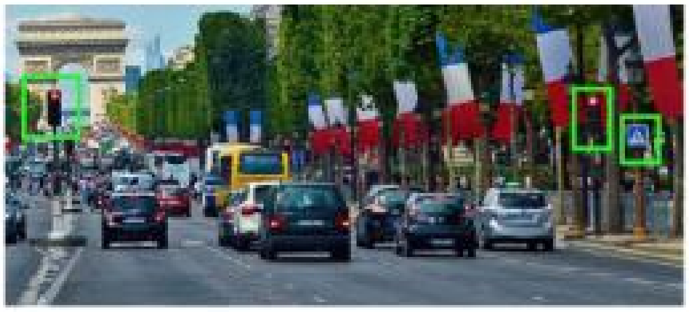
1. To help you practice strategies for machine learning, this week we'll present another scenario and ask how you would act. We think this "simulator" of working in a machine learning project will give you an idea of what leading a machine learning project could be like!

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, 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.

Suppose that you came from working with a project for human detection in city parks, so you know that detecting humans in diverse environments can be a difficult problem. What is the first thing you do? Assume each of the steps below would take about an equal amount of time (a few days).

- ( ) Start by solving pedestrian detection, since you already have the experience to do this.
- (•) Train a basic model and proceed with error analysis.
- ( ) Spend a few days collecting more data to determine how hard it will be to include more pedestrians in your dataset.
- ( ) Leave aside the pedestrian detection, to move faster and then later solve the pedestrian problem alone.

Expand

Correct

Correct. As discussed in the lecture, it is better to create your first system quickly and then iterate.

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.

0 / 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?

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

Expand

Incorrect

The logistic loss doesn't work here since the $y^{(i)}$ and $\hat{y}^{(i)}$ are vectors.

3. You are carrying out error analysis and counting up what errors the algorithm makes. Which of these datasets do you think you should manually go through and carefully examine, one image at a time?

1 / 1 point

- ○ 10,000 images on which the algorithm made a mistake
- ◉ 500 images on which the algorithm made a mistake
- ○ 10,000 randomly chosen images
- ○ 500 randomly chosen images

Expand

Correct

Focus on images that the algorithm got wrong. Also, 500 is enough to give you a good initial sense of the error statistics. There's probably no need to look at 10,000, which will take a long time.

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, 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?

- ○ True
- ◉ False

Expand

Correct

As seen in the lecture on multi-task learning, you can compute the cost such that it is not influenced by the fact that some entries haven't been labeled.

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 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.
- ◉ 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.

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)	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, a friend thinks that the training data distribution is much harder than the dev/test distribution. What do you think?

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

Expand

Correct

Correct. Since the training-dev error is higher than the dev and test errors, the dev/test distribution is probably "easier" than the training distribution.

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 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.
- ○ First start with the sources of error that are least costly to fix.
- ○ True because it is greater than the other error categories added together $8.0 > 4.1 + 2.2 + 1.0$.
- ◉ 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 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 how much this windshield wiper could worsen performance in the worst case.
- ○ 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 the minimum amount this windshield wiper could improve performance.

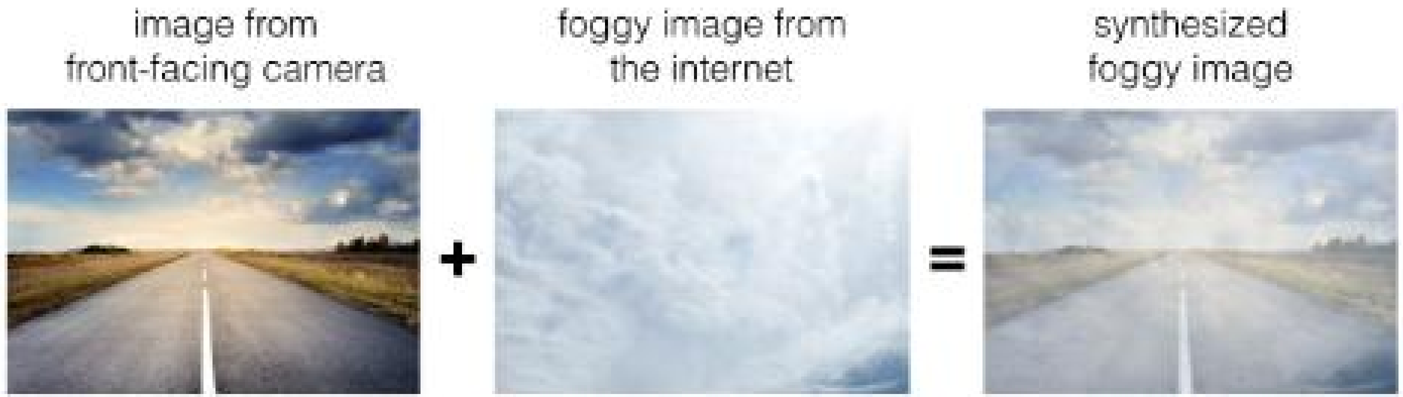
Expand

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 do you agree with?

- ○ It is irrelevant how the resulting foggy images are perceived by the human eye, the most important thing is that they are correctly synthesized.
- ○ With this technique, we duplicate the size of the training set by synthesizing a new foggy image for each image in the training set.
- ○ If used, the synthetic data should be added to the training/dev/test sets in equal proportions.
- ◉ If used, the synthetic data should be added to the training set.

Expand

Correct

Yes. The synthetic data can help to train the model to get better performance at the dev set, but shouldn't be added to the dev or test sets because they don't represent our target in a completely accurate way.

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 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.
- ○ False, the test set shouldn't be changed since we want to know how the model performs in real data.

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. 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?

- [x] 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?

1 / 1 point

- [x] 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 red and green lights, you have been using this approach:

1 / 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

Expand

Correct
Yes. (A) is an end-to-end approach as it maps directly the input (x) to the output (y).

15. To recognize a stop sign you use the following approach:

0 / 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 are available models which we can use to transfer knowledge.
- ○ 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.

Expand

Incorrect

No. In this case, it is more promising to use an end-to-end approach.