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!

   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 image below 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 theinternet, which could be helpful for training even if the distribution of internet data is not the same.

   You are getting started with 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 checking how well humans perform at these tasks.
   - ◉ Spend a few days training a basic model and see what mistakes it makes.
   - ◯ Spend a few days getting the internet data, so that you understand better what data is available.

   ✓ **Correct**
   As discussed in lectures, machine learning is an iterative process. Training a basic model and analyzing its errors will 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, **which of the following gives you the most appropriate activation function?**

   - ◯ ReLU
   - ◯ Softmax
   - ◯ Linear
   - ◉ Sigmoid

   ✓ **Correct**
   Sigmoid is appropriate because it outputs values between 0 and 1, representing the probability that each individual object (road sign or traffic light) is present in an image. This allows for multiple objects to be detected in a single image.

3. **True or False:** When trying to determine what strategy to implement to improve the performance of a model, you manually check all images of the training set where the algorithm was successful.

- (•) False
- ( ) True

Correct
The training set is typically very large, and manually checking all the successful images would be time-consuming and inefficient. It's more effective to focus on the errors in the dev set to understand where the model needs improvement.

4. After working on the data for several weeks, your team ends up with the following data:

- 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.

True or False: In multi-task learning, if some examples have missing labels (for example: $\begin{bmatrix} 0 \\ ? \\ 1 \\ 1 \\ ? \end{bmatrix}$), the learning algorithm **cannot** use those examples.

- ( ) True
- (•) False

Correct
As seen in the lecture on multi-task learning, you can compute the cost function in a way that it is not affected by missing labels. The algorithm can still learn from the available labels in the example.

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.

Which of the following are true about the train/dev/test split?

- [ ] The train, dev and test sets must come from the same distribution.
- [x] The dev and test sets must come from the same distribution.

Correct
This is required to make sure that the dev and test sets accurately reflect the data you ultimately want the model to perform well on.

- [ ] The dev and test sets must contain some images from the internet.
- [x] The dev and test set must come from the front-facing camera.

Correct
This is the distribution you care about most; thus you should use this as a target.

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:

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.

True/False: Should the team's highest priority be to bring more foggy pictures into the training set to address the 8.0% of errors in that category?

- (●) 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.
- ( ) 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.

Correct
You should consider the trade-off 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.

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 one of the following statements do you agree with?

- (●) 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 will improve performance.
- ( ) 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 could worsen performance in the worst case.

Correct
You will probably not improve performance by more than 2.2% by solving the raindrops problem.

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:

Which one 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 set.
- ○ If used, the synthetic data should be added to the training/dev/test sets in equal proportions.

✓ **Correct**
The synthetic data can help train the model to get better performance on the dev set, but it 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.

True or False: You need to correct the labels of the test set so that the 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 that they aren't severely affected by the difference in distributions.

- ○ False, the test set shouldn't be changed since we want to know how the model performs with uncorrected or original 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. The labels in the training set should be fixed only in case of a systematic error.

✗ **Incorrect**
Due to the robust nature of deep learning models, we don't need to have the same distribution between train, dev, and test sets.

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.

True or False: 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

✓ **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.

    True or False: This case could benefit from using multi-task learning.

    - (●) True
    - ( ) False

    Correct
    Multi-task learning is suitable here due to the shared high-level features among the required road signs.

14. You want to recognize red and green lights in images. You have two approaches:

    - **Approach 1**: Input an image (x) into a neural network that directly predicts whether a red or green light is present (y).
    - **Approach 2**: First, detect the traffic light in the image (if any). Then, determine the color of the illuminated lamp.

    Which approach is a better example of an end-to-end approach?

    - ( ) Approach 2
    - (●) Approach 1

    Correct
    Approach 1 directly maps the input (x) to the output (y) in a single step, which is the definition of an end-to-end approach.

15. **True or False:** An end-to-end approach doesn't require that we hand-design useful features; it only requires a large enough model.

    - (●) True
    - ( ) False

    Correct
    This is a key characteristic of end-to-end deep learning models, which learn features directly from the data.