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 an idea 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, this 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, 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 some time searching the internet for the data most similar to the conditions you expect on production.
- ◉ Train a basic model and do error analysis.
- ◯ Invest a few days in thinking on potential difficulties, and then some more days brainstorming about possible solutions, before training any model.
- ◯ 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.

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

For the output layer, **which of the following gives you the most appropriate activation function?**

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

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. 1 point

- ◯ True
- ◉ False

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.

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.

- ◉ 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. 1 point

How should you split the dataset into train/dev/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 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.
- ◯ 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.

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

Dataset:	Contains:	Error of the algorithm:
Training	940,000 images randomly picked from (900,000 internet images +	8.8%

	60,000 car's front-facing camera images)	
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?

- [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.
- [ ] You have a large variance problem because your training error is quite higher than the human-level error.
- [x] You have a large avoidable-bias problem because your training error is quite a bit 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.

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

Human-level error on this task is approximately 0.5%. (Bayes error is the lowest possible error rate for a task. Human-level error is a good estimation of Bayes error.)

True or False: Based on this, the Bayes error for the car camera images (Dev/Test) is higher than the Bayes error for the mixed internet/car images (Training).

- ○ True
- ◉ 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	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.
- ◉ 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.
- ○ 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.

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.

True or False: From this table, you can conclude that if you fix the incorrectly labeled data you will reduce the overall dev set error to 11.2%.

- ◉ 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

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.
- ○ 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.
- ○ With this technique, we duplicate the size of the training set by synthesizing a new foggy image for each image in the training set.

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

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 should be changed, but also the train set to keep the same distribution between the train, dev, and test sets.
- ◉ False, the test set shouldn't be changed since we want to know how the model performs with uncorrected or original data.
- ○ 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.

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 point

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

13. Another colleague wants to use microphones placed outside the car to better hear if there are other vehicles around you. For example, if there is a police vehicle behind you, you would be able to hear their siren. However, they don't have much data to train this audio system. How can you help? 1 point

- ◉ Neither transfer learning nor multi-task learning seems promising.
- ○ Either transfer learning or multi-task learning could help our colleague get going faster.
- ○ Transfer learning from your vision dataset could help your colleague get going faster. Multi-task learning seems significantly less promising.
- ○ Multi-task learning from your vision dataset could help your colleague get going faster. Transfer learning seems significantly less promising.

14. You're building a system to recognize stop signs. Your approach is: 1 point

First, a neural network predicts bounding box coordinates around potential traffic signs in an image. **Second**, a separate neural network determines if each predicted sign is a stop sign.

Is this an example of multi-task learning?

- ◉ False
- ○ True

15. To recognize a stop sign, you use the following approach: 1 point

First, localize any traffic sign in an image. **After that**, 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 a large amount of data.
- ○ The problem has a high Bayes error.
- ◉ There is not enough data to train a big neural network.
- ○ There are available models which we can use to transfer knowledge.