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 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).

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

74:45

Expand

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?

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

74:44

Expand

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 point

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

74:44

Expand

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.

When using a non fully labeled image such as $y^{(i)} = \begin{bmatrix} 0 \\ ? \\ 1 \\ ? \\ 1 \end{bmatrix}$, which of the following strategies is most appropriate to calculate the loss function to train as a multi-task learning problem?

- It is not possible to use non fully labeled images if we train as a multi-task learning problem.
- Make the missing entries equal to 1.
- Make the missing entries equal to 0.
- Calculate the loss as $\sum \mathcal{L}(\hat{y}_j^{(i)}, y_j^{(i)})$ where the sum goes over all the know components of $y^{(i)}$.

74:43

Expand

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.
- 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 20,000 images from your car's front-facing camera. The 80,000 remaining images will be split equally in dev and test sets.

74:43

Expand

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 + 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?

- You have a large data-mismatch problem.
- You have a high variance problem.
- You have a high bias.
- The size of the train-dev set is too high.

74:42

Expand

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)	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?

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

74:41

Expand

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 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.
- You should prioritize getting more foggy pictures since that will be easier to solve.

74:41

Expand

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.

You find out that there is an anti-reflective film guarantee to eliminate the sun reflection, but it is quite costly. Which of the following gives the best description of what the investment in the film can do to the model?

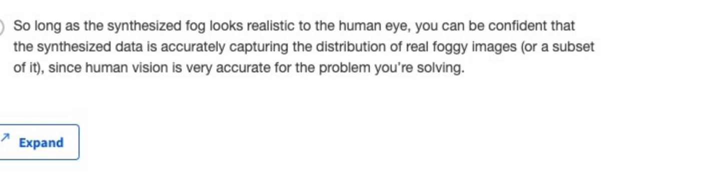
- The overall test set error will be reduced by at most 7.2%.
- The film will reduce at least 7.2% of the dev set error.
- The film will reduce the dev set error with 7.2% at the most.

74:40

Expand

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 of the following statements do you agree with?

- 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.
- 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.
- 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.

74:40

Expand

11. After working further on the problem, you've decided to correct the incorrectly labeled data on the dev set. Which of these statements do you agree with? (Check all that apply).

1 point

- [ ] You should not correct the incorrectly labeled data in the test set, so that the dev and test sets continue to come from the same distribution.
- [ ] You should correct incorrectly labeled data in the training set as well so as to avoid your training set now being even more different from your dev set.
- [ ] You do not necessarily need to fix the incorrectly labeled data in the training set, because it's okay for the training set distribution to differ from the dev and test sets. Note that it is important that the dev set and test set have the same distribution.
- [ ] You should also correct the incorrectly labeled data in the test set, so that the dev and test sets continue to come from the same distribution.

74:39

Expand

12. Your client asks you to add the capability to detect dogs that may be crossing the road to the system. He can provide a relatively small set containing dogs. Which of the following do you agree most with?

1 point

- You will have to re-train the whole model now including the dogs' data.
- Using pre-trained weights can severely hinder the ability of the model to detect dogs since they have too many learned features.
- You can use weights pre-trained on the original data, and fine-tune with the data now including the dogs.
- You should train a single new model for the dogs' task, and leave the previous model as it is.

74:39

Expand

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 point

- False
- True

74:39

Expand

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 point

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

74:38

Expand

15. Consider the following two approaches, A and B:

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).

· **(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.

Approach A tends to be more promising than approach B if you have a ________ (fill in the blank).

- Multi-task learning problem.
- Large training set
- Problem with a high Bayes error.
- Large bias problem.

74:37

Expand