Your grade: 100%

Your latest: **100%** • Your highest: **100%** • To pass you need at least 80%. We keep your highest score.

Next item →

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.

$$y^{(i)} = \begin{array}{c l} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} & \begin{array}{l} \text{"stop sign"} \\ \text{"pedestrian crossing sign"} \\ \text{"construction ahead sign"} \\ \text{"red traffic light"} \\ \text{"green traffic light"} \end{array} \end{array}$$

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 theinternet, 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).

- 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.
- Start by solving pedestrian detection, since you already have the experience to do this.
- Train a basic model and proceed with error analysis.

Correct
It's most efficient to create a basic system and iterate based on what the errors reveal.

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 / 1 point

True or False: For the output layer, a softmax activation would be a good choice because this is a multi-task learning problem.

- False
- True

Correct
Softmax would be a good choice if one and only one of the possibilities (stop sign, speed bump, pedestrian crossing, green light and red light) was present in each image. However, in this case, multiple objects can be present in a single image. Therefore, a sigmoid activation for each output node is more appropriate.

3. You are carrying out error analysis and counting up what errors the algorithm makes.

1 / 1 point

Which of these datasets do you think you should manually go through and carefully examine, one image at a time?

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

✓ Correct
Focusing on images that the algorithm got wrong helps you understand its weaknesses. 500 is a reasonable number to start with to get a good initial understanding of the error patterns.

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.

True or False: Because this is a multi-task learning problem, when an image is not fully labeled (for example: $\begin{pmatrix} 0 \\ ? \\ ? \\ 1 \\ 0 \end{pmatrix}$) you can use it if you ignore those entries when calculating the loss function.

- (●) True
- ( ) False

✓ Correct
You *can* use partially labeled images. By ignoring the missing labels and calculating the loss based on the known labels, you can still effectively train the model.

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 / 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 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.

✓ Correct
Yes. 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 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?**

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

✓ Correct
The large difference between the training error and human-level error indicates that there's room for improvement in the model's ability to fit the data

The large difference between the training error and human-level error indicates that there's room for improvement in the model's ability to fit the data.

- [ ] You have a large variance problem because your training error is quite higher than the human-level error.
- [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**
> The significant difference between the training-dev error and dev error suggests that the model is performing better on the training distribution than on the "real-world" data from your car's camera.

- [ ] 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 / 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

> **Correct**
> The Dev/Test errors (1.3% and 1.1%) are lower than the Training/Training-Dev errors (2% and 2.3%). This indicates the car camera images are likely inherently easier for the algorithm to learn, thus having a lower Bayes error.

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 / 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.

True or False: 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

> **Correct**
> While the sum of other errors is greater, focusing solely on the numbers ignores crucial factors like the cost and accessibility of acquiring more images with partially occluded.

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 one 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 the minimum amount this windshield wiper could 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 how much this windshield wiper will improve performance.

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: **1 / 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.

○ 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.

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. **1 / 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.

Correct
To successfully train a model, the dev set and test set should come from the same distribution. Also, 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. 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 refer to it as an orange light; however, we will 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 / 1 point**

What do you tell your colleague?

◉ She should try using weights pre-trained on your dataset and fine-tuning further with the yellow-light dataset.

○ Recommend that she try multi-task learning instead of transfer learning using all the data.

○ 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.

from "swamping" the yellow lights dataset.

- [ ] You cannot help her because the distribution of data you have is different from her's and is also lacking the yellow label.

> **Correct**
> You have trained your model on a large dataset, and she has a small dataset. Although your labels are different, the parameters of your model have been trained to recognize many characteristics of road and traffic images, which will be useful for her problem. This is a perfect case for transfer learning; she can start with a model with the same architecture as yours, change what is after the last hidden layer, and initialize it with your trained parameters.

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. **1 / 1 point**

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

- [ ] False
- [x] True

> **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: **1 / 1 point**

- **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?

- [x] Approach 1
- [ ] Approach 2

> **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. To recognize a stop sign, you use the following approach: **1 / 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**.

- [ ] The problem has a high Bayes error.
- [ ] There is a large amount of data.
- [x] There is not enough data to train a big neural network.
- [ ] There are available models which we can use to transfer knowledge.

> **Correct**
> When data is limited, a two-step approach can be more effective than training a large end-to-end model.