Congratulations! You passed!

Grade received 80% Latest Submission Grade 80% To pass 80% or higher

Go to 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!

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.

Suppose that you same from working on this 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.
- Train a basic model and proceed with error analysis.
- 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.

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. 1 / 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?

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

Expand

Incorrect
Notice that $\hat{y}_j^{(i)}$ indicates the component i of the example x.

3. When trying to determine what strategy to implement to improve the performance of a model, we manually check all images of the training set where the algorithm was successful. True/False? 1 / 1 point

- True
- False

Expand

Correct
Correct. This set should be too large to manually check all the images. It is better to focus on the images that the algorithm got wrong from the dev set. Also, choose a large enough subset that we can manually check.

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, $y^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ means the image contains a stop sign and a red traffic light.

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

True/False?

- True
- False

Expand

Correct
Correct. We can't use the components of the labels that are missing but we can use the ones we have to 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. Which of the following are true about the train-dev held out? 1 / 1 point

- The dev and test set must come from the front-facing camera.
- The dev and test sets must come from the same distribution.
  - Correct
    Correct. This is required to aim the target where we want to be.
- The train, dev, and test must come from the same distribution.
- The dev and test sets must contain some images from the internet.
  - This should not be selected
    The dev and test set should come from the distribution that we care about. In this case, that is the front-facing camera.

Expand

Incorrect
You didn't select all the correct answers

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.

- 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.
- You have a large avoidable bias problem because your training error is quite a lot higher than the human level error.
  - Correct
    Your algorithm overfits the dev set because the error of the dev and test sets are very close.
- 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	1.5%

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 you conclude that the Bayes error for the dev/test distribution is probably higher than for the train distribution. True/False?

- True
- False

Expand

Correct

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 mistakes). For example, about 7.2/15.3 = 47% of your errors are due to partially occluded elements.

You shouldn't invest all your efforts to get more images with partially occluded elements since 1 × 3.0 + 1.0 + 4.1 = 7.2. True/False?

- True
- False

Expand

Correct
Correct. These kinds of arguments don't help us to decide on the strategy to follow. Other factors should be used, such as the tradeoff between the cost of getting new images and the improvement of the system performance.

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 / 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 mistakes). For example, about 7.2/15.3 = 47% of your errors are due to partially occluded elements in the image.

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

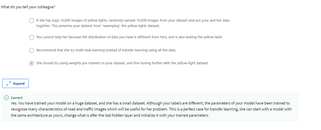
- False
- True

Expand

Correct
Correct. The 4.1 only gives us a ceiling of how much the error can be improved by fixing the labels.

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

image from front-facing camera + foggy image from the internet = synthesized foggy image



We can't use this data since they have a different distribution from the ones we used (internet and front-facing camera). True/False?

- True
- False

Expand

Correct
Correct. The new synthesized images are added to the training set and as long as they look realistic to the human eye this will be useful data to train the model.

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 or 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 a practical rule, we most keep dev and test with the same distribution. And the labels of 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. 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 call it an orange light rather than a yellow light, we'll use the US convention of calling it yellow.) Images containing yellow lights are quite rare, and the dream team 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?

- 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.
- You cannot help her because the distribution of data you have is different from hers, and is also lacking the yellow label.
- Recommend that she first multi-task learning instead of transfer learning using all the data.
- She should try using weights pre-trained on your dataset, and fine-tuning further with the yellow light dataset.

Expand

Correct
Yes. You have trained your model on a huge 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's 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 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 / 1 point

- True
- False

Expand

Correct
Correct. When using transfer learning we can remove the last layer. That is one of the aspects that is different from a binary classification problem.

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 / 1 point

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

Expand

Incorrect
An end-to-end approach avoids the use of hand-designed components.

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

First, we locate 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 is not enough data to train a big neural network.
- The problem has a High Bayes error.
- There are available models which we can use to transfer knowledge.
- There is a large amount of data.

Expand

Correct
Correct. This might be the most important factor when deciding whether to use an end-to-end approach.