Congratulations! You passed!

Grade received 93.33%

Latest Submission Grade 93.33%

To pass 80% or higher

Retake the assignment in **7h 6m**

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 a task 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, the above 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, that could be helpful for training even if the distribution of internet data is not the same.

You are just getting started on 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 checking what is human-level performance for these tasks so that you can get an accurate estimate of Bayes error.
- ○ Spend a few days getting the internet data, so that you understand better what data is available.
- ○ 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 training a basic model and see what mistakes it makes.

Expand

Correct

As discussed in lecture, applied ML is a highly iterative process. If you train a basic model and carry out error analysis (see what mistakes it makes) it will help 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, a softmax activation would be a good choice for the output layer because this is a multi-task learning problem. True/False?

1 / 1 point

- (●) False
- ( ) True

Expand

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.

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

Because this is a multi-task learning problem, you need to have all your $y^{(i)}$ vectors fully labeled. If one example is equal to $\begin{bmatrix} 0 \\ ? \\ 1 \\ 1 \\ ? \end{bmatrix}$ then the learning algorithm will not be able to use that example. True/False?

- (●) False
- ( ) True

Correct
As seen in the lecture on multi-task learning, you can compute the cost such that it is not influenced by the fact that some entries haven't been labeled.

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. The best way to split the data is using the 900,000 internet images to train, and divide the 100,000 images from your car's front-facing camera between dev and test sets. True/False? **1 / 1 point**

- (•) False
- ( ) True

Expand

Correct
Correct. 100,000 images are too many to use in dev and test. A better distribution would be to use 80,000 of those images to train, and split the rest between dev and test.

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).

- [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**

- [ ] Your algorithm overfits the dev set because the error of the dev and test sets are very close.

- [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**

- [ ] Your algorithm overfits the dev set because the error of the dev and test sets are very close.

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

  ✓ **Correct**

- [ ] 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 variance problem because your training error is quite higher than the human-level error.

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	1.1%

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 harder than the dev/test distribution. What do you think?

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

Expand

Correct
Correct. Since the training-dev error is higher than the dev and test errors, the dev/test distribution is probably "easier" than the training distribution.

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	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?

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.
- ( ) You should prioritize getting more foggy pictures since that will be easier to solve.
- ( ) Since 8.2 > 4.1 + 2.0 + 1.0, the priority should be to get more images with partially occluded elements.
- ( ) 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.

Expand

Correct

Correct. You should consider the tradeoff 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. 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 of the following statements do you agree with?

- 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 the maximum 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.

Expand

Correct

Yes. You will probably not improve performance by more than 2.2% by solving the raindrops problem. If your dataset was infinitely big, 2.2% would be a perfect estimate of the improvement you can achieve by purchasing a specially designed windshield wiper that removes the raindrops.

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 of the following do you agree with?

- ○ 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.
- ○ 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.

Expand

Correct

Yes. The synthetic data can help to train the model to get better performance at the dev set, but 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. **0 / 1 point**

You have to correct the labels of the test so 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?

- ○ True, as pointed out, we must keep dev and test with the same distribution. And the labels at training should be fixed only in case of a systematic error.
- ◉ 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 in real data.

Expand

⊗ **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.

1 / 1 point

But 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?

- ○ True
- ◉ False

Expand

Correct

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. 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 to train this audio system. How can you help?

1 / 1 point

- ○ Multi-task learning from your vision dataset could help your colleague get going faster. Transfer learning seems significantly less promising.
- ◉ 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.

Expand

Correct

Yes. The problem he is trying to solve is quite different from yours. The different dataset structures make it probably impossible to use transfer learning or multi-task learning.

14. **To recognize a stop sign you use the following approach:** First, you use a neural network to predict bounding box co-ordinates around all traffic signs (if any) within an input image. You then pass the results to a different neural network to determine if the predicted traffic signs (if any) are a stop sign or not. We are using multi-task learning. True/False? 1 / 1 point

- (●) False
- ( ) True

Expand

Correct

Correct. Multi-task learning is about joining several tasks that can benefit from each other. Since there are 2 different neural networks being used here that do not share weights (i.e. structure), this problem has 2 single task learning neural networks and not a multi-task learning setup.

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

First, we localize 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.

- ○ The problem has a high Bayes error.
- ○ There are available models which we can use to transfer knowledge.
- ◉ There is not enough data to train a big neural network.
- ○ 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.