

Autonomous driving (case study)

Submit your assignment

ETA DATE Jan 6, 1:00 AM CSTATTORNEY 3 days & hours

Review grade

10 PASS 10% or higher

100%

View feedback

Try again

See how your progress compares

✓ Congratulations! You passed!

10 PASS 10% or higher

Keep Learning

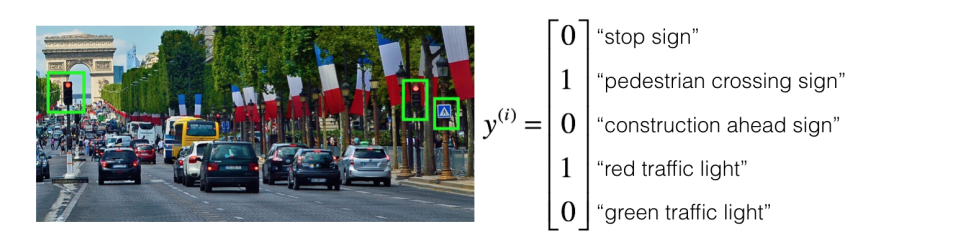
100%

Autonomous driving (case study)

LATEST SUBMISSION GRADE
100%

1. To help you practice strategies for machine learning, in this week we present another scenario and ask how you would approach this "toy problem" of working in a machine learning project and give a list of what building 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, the above image contains a pedestrian crossing sign and red traffic lights.



Your 100,000 labeled images are taken using the front-facing camera of your car. This is also the distribution of data you can trust about being on- or off-the-road. You might also get a much larger dataset of the internet, but that's a bit helpful for training even if the distribution of internet data is not the same.

You are going to build 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 training a basic model and see what mistakes it makes.
- ☐ Spend a few days collecting more data using the front-facing camera of your car, to better understand how much data per car/time you can collect.
- ☐ Spend a few days checking what is human-level performance for these tasks so that you can get an accurate estimate of expert error.
- ☐ Spend a few days getting the internet data, so that you understand better what data is available.

✓ **Correct**
As discussed in lecture, applying ML is a highly iterative process. If you have a basic model and carry out error analysis on what mistakes it makes, it will help you to more precisely direct them.

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 1000 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 multiclass learning problem. True/False?

- ☐ True
- ☒ False

✓ **Correct**
Softmax would be a good choice if one and only one of the possibilities (stop sign, speed bump, pedestrian crossing, green light or red light) was present in each image.

3. You are carrying out error analysis and counting up what errors the algorithm makes. Which of these datasets do you think you should randomly go through and carefully examine one image at a time?

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

✓ **Correct**
Reviewing images that the algorithm got wrong (500, 1000 is enough to give you a good initial sense of the error statistics). Thereby, probably to look at 10,000, which will take a long time.

4. After working on the data for several weeks, your team ends up with the following data:

- 100,000 labeled images taken using the front-facing camera of your car.
- 100,000 labeled images of roads downloaded from the internet.
- Each image's labels properly indicate the presence of any specific road sign and traffic signals or combinations of them. For example, $y^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$ means the image contains a stop sign and a red traffic light.

Because this is a multiclass learning problem, you need to have all your $y^{(1)}$ 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?

- ☐ True
- ☒ False

✓ **Correct**
As seen in the lecture on multiclass 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 can access 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?

- ☐ Mix all the 100,000 images with the 100,000 images you found online. Shuffle everything. Split the 1,000,000 images dataset into 100,000 for the training set, 10,000 for the dev set and 10,000 for the test set.
- ☒ Choose the training set to be the 100,000 images from the internet along with 90,000 images from your car's front-facing camera. The 10,000 remaining images will be split equally to dev and test sets.
- ☐ Mix all the 100,000 images with the 100,000 images you found online. Shuffle everything. Split the 1,000,000 images dataset into 100,000 for the training set, 10,000 for the dev set and 10,000 for the test set.
- ☐ Choose the training set to be the 100,000 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 to dev and test sets.

✓ **Correct**
As we saw in lecture, 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 "hard" data to avoid having a data research problem.

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

Dataset	Contents	Error of the algorithm
Training	94,000 images randomly picked from 100,000 internet images + 60,000 car's front-facing camera images	8.4%
Training	10,000 images randomly picked from 100,000 internet images + 10,000 car's front-facing camera images	9.7%
Dev	10,000 images from your car's front-facing camera	14.2%
Test	10,000 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 data-mismatch problem because your model does a lot better on the training data set than on the dev set.
- ☒ You have a large available 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 but that they were seen before.
- ☐ Your algorithm overfits the dev set because the error of the dev and test sets are very close.

7. Based on table from the previous question, a third reads that the training data distribution is much noisier than the dev/test distribution. What do you think?

- ☐ Your friend is right. If a 1.8% error for the training data distribution is probably lower than for the dev/test distribution.
- ☐ Your friend is wrong. If a 1.8% error for the training data distribution is probably higher than for the dev/test distribution.
- ☒ There's insufficient information to tell if your friend is right or wrong.

✓ **Correct**
The algorithm does better on the distribution of data it was trained on. But you don't know if it's because it trained on that distribution or if it's really a bias. To get a better sense, measure human-level error separately on both distributions.

8. You decide to focus on the dev set and check by hand what are the errors due to. Here is a table summarizing your observations:

Source of error	Fraction
Errors due to incorrectly labeled data	4.7%
Errors due to blurry pictures	8.2%
Errors due to things such as your car's front-facing camera	2.2%
Errors due to other causes	1.9%

In this table, 4.7%, 8.2%, etc. are a fraction of the total dev set (not just examples your algorithm misclassified). For example, about 0.0175 = 1.75% of your errors are due to blurry pictures.

The results from this analysis tell us that the most significant problem should be to bring more blurry pictures into the training set so as to address the 8.2% of errors in that category. True/False?

Additional note: there are subtle concerns to consider with this question, and you may find arguments for why some answers are also correct or incorrect. We recommend that you spend time reading the feedback for this task to understand what issues that you want to consider when you are building your own machine learning project.

- ☐ True because it is the largest category of errors. We should always prioritize the largest category of error as this will make the best use of the team's time.
- ☐ True because it's greater than the other error categories added together (8.2% + 4.7% + 2.1%).
- ☒ False because it depends on how easy it is to add blurry data. If blurry data is very hard and costly to collect, it might not be worth the team's effort.
- ☐ First start with the sources of error that are least costly to fix.

✓ **Correct**
Correct feedback. This is the correct answer. 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 waterdrops on the front-facing camera. Based on the table from the previous question, which of the following statements do you agree with?

- ☒ 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 the minimum amount this windshield wiper could improve performance.
- ☐ 2.2% would be a reasonable estimate of how much this windshield wiper will improve performance.
- ☐ 2.2% would be a reasonable estimate of how much this windshield wiper could worsen performance in the worst case.

✓ **Correct**
Yes. This is a well-posed question. It's important to be clear that 2.2% by adding the windshield wiper to your dataset was effectively 0.2% would be a perfect estimate of the improvement you can achieve by purchasing a specially designed windshield wiper that removes this raindrops.

10. You decide to use data augmentation to address blurry images. You find 1,000 pictures of fog off the internet, and "add" them to clean images to synthesize blurry images, like this:



Which of the following statements do you agree with?

- ☐ There is a little risk of overfitting to the 1,000 pictures of fog so long as you are combining it with a much larger set of clean images from your car.
- ☒ Adding synthesized images that look like real foggy pictures taken from the front-facing camera of your car to training dataset will help the model generalize better and will introduce available 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 this problem you're solving.

✓ **Correct**
Yes. If the synthesized images look realistic, then the model will just see them as if you had added useful data to synthesize road signs and traffic signals in a foggy weather. Not very helpful.

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).

- ☒ 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.
- ☒ Yes. Because you want to make sure that your dev and test data come from the same distribution for your algorithm to make your model's results development process as robust.
- ☐ You should correct incorrectly labeled data in the training set as well as so as to avoid your training set now being even more different from your dev set.
- ☐ 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 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.
- ☒ Yes. True, deep learning algorithms are quite robust to having slightly different train and dev distributions.

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 yellow traffic lights. Their dataset is of a single light after their yellow light, with set 10% of the dataset consisting of yellow traffic lights. Images contain yellow lights are quite rare, and she doesn't have enough data to build a good model. She hopes you can help her by using transfer learning.

What do you tell your colleague?

- ☒ She should try using weights pre-trained on your dataset, and fine-tune further with the yellow-light dataset.
- ☐ If she has less than 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 "overfitting" to the yellow light dataset.
- ☐ You cannot help her because the distribution of data you have is different from hers, and is also lacking the yellow light.
- ☐ Recommend that she try multi-task learning instead of transfer learning using all the data.

✓ **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 learned to recognize thousands of red and green 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 it sees after the last hidden layer and initialize it with your trained parameters.

13. Another colleague wants to use microphones placed outside the car to better hear if someone 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 hear with a microphone. How do you help?

- ☐ 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.
- ☐ Other transfer learning or multi-task learning could help, but your colleague got going faster.
- ☒ Neither transfer learning nor multi-task learning seems promising.

✓ **Correct**
Yes. The problem he's 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 red and green lights, you have been using this approach:

- ☒ Map each image pixel to a neural network and have it directly learn a mapping to make a prediction as to whether there's a red light or not a green light (x).

A teammate proposes a different, two-step approach:

- ☒ If it's too noisy, your approach, you could first do about the traffic light in the image (if only, then 1) determine the color of the fluorescent lamp in the traffic light.

Between these two, Approach 1 is more of an end-to-end approach because it has distinct steps for the input and the output and True/False?

- ☐ True
- ☒ False

✓ **Correct**
Yes. (X) is an end-to-end approach as it maps directly the input (x) to the output (y).

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

- ☒ Large amount of data.
- ☐ Multi-task learning problem.
- ☐ Large bias problem.
- ☐ Problem with a high Bayes error.

✓ **Correct**
Yes. In many fields, it has been observed that end-to-end learning works better in practice, but requires a large amount of data.