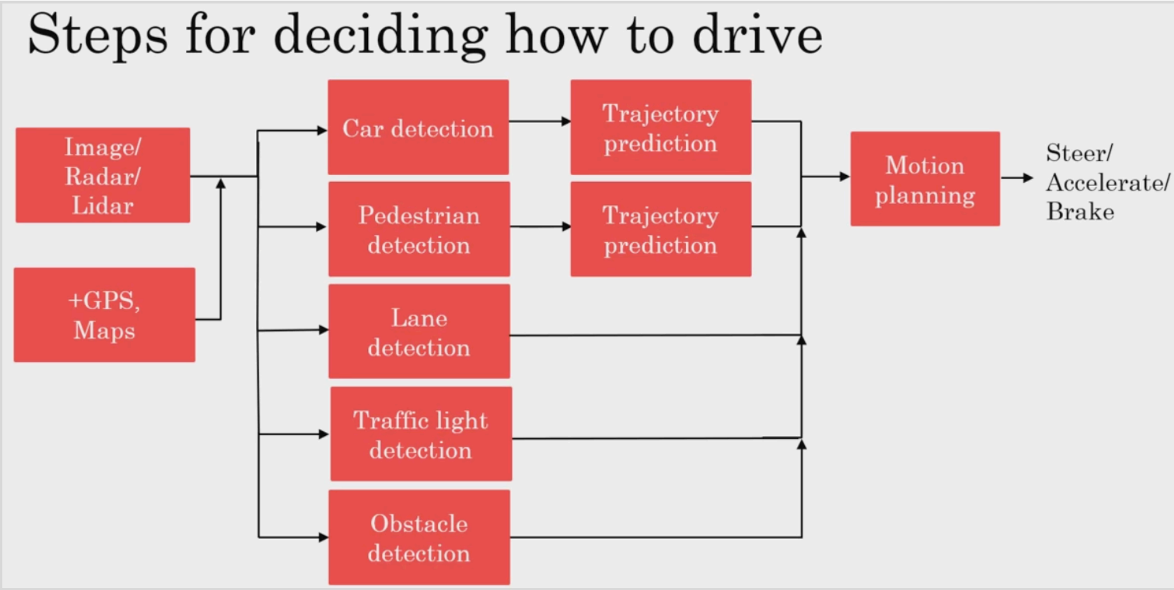
# Autonomous Driving (Case Study)




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!

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.

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

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, 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).

- [x] Spend a few days training a basic model and see what mistakes it makes.
- [ ] 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.

Note: 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.

Suppose that you use a sigmoid function for the output layer, and the output yhat has shape (5, 1). Which of the following best describes the cost function?

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

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?

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

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.
- 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, when an image is not fully labeled (for example: $\begin{pmatrix}0\\?\\?\\1\\0\end{pmatrix}$) we can use it if we ignore those entries when calculating the loss function. True/False?

- [x] True
- [ ] False

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/test split?

- [ ] The dev and test sets must contain some images from the internet.
- [x] The dev and test sets must come from the same distribution.
- [x] The dev and test set must come from the front-facing camera.
- [ ] The train, dev, and test must come from the same distribution.

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

| **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 model is not generalizing well to data from the same training distribution but that it has never seen before
- [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.
- [ ] Your algorithm overfits the dev set because the error of the dev and test sets are very close.
- [ ] You have a large variance problem because your training error is quite higher than the human-level error.
- [x] You have a large avoidable-bias problem because your training error is quite a bit higher than the human-level error.

7. Assume you’ve finally chosen the following split between the data:

| **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?

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

Note: 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:

| | |
|---|---|
| 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?

- [x] 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 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.
- [ ] 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.

Note: You should consider the tradeoff between the data accessibility and potential improvement of your model trained on this additional data.

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:

| | |
|---|---|
| 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% |

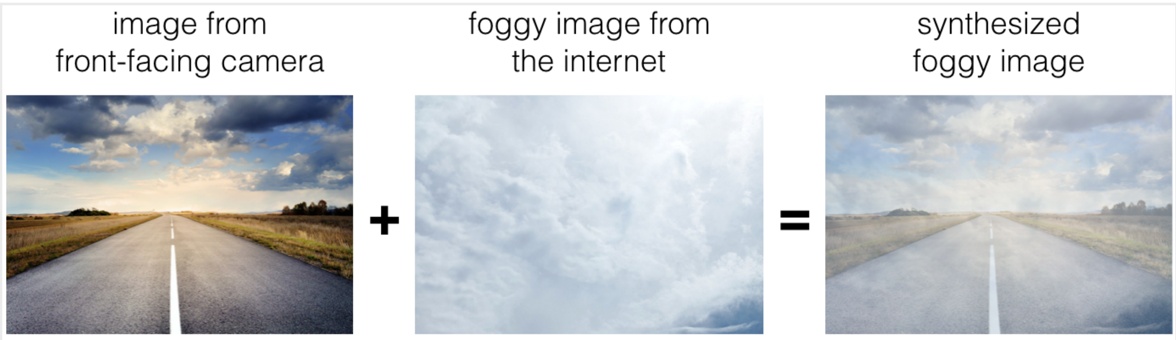
In this table, 4.1%, 7.2%, etc. are a fraction of the total dev set (not just examples of your algorithm mislabeled). For example, about 7.2/15.3 = 47% of your errors are due to partially occluded elements.

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

Note: Remember that this 7.2% gives us an estimate for the ceiling of how much the error can be reduced when the cause is fixed.

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:




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.
- [x] 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.

Note: If the synthesized images look realistic, then the model will just see them as if you had added useful data to identify road signs and traffic signals in foggy weather. It will very likely help.

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. Which of the following is a necessary step to take?

- [ ] Use a correctly labeled version and an incorrectly labeled version to make the model more robust.
- [ ] Create a train-dev set to estimate how many incorrectly labeled examples are in the train set.
- [ ] Correct the labels of the train set.
- [x] Correct the labels of the test set.

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?

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

Note: Since your model has learned useful low-level features to tackle the new task we can conserve those by using the pre-trained weights.

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?

- [x] False
- [ ] True

Note: 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?

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

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

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