Structuring Machine Learning Projects

Week 2: 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 an idea of what leading a machine learning project could be like! [..,omitted...] You are getting started with 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).

Answer: Train a basic model and do error analysis.

Comment: Applied ML is highly iterative. Having a basic model to do an error analysis can point you in the most promising directions with a lot of certainties.

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, which of the following gives you the most appropriate activation function?

Answer: Sigmoid.

Comment: This works well since the output would be valued between 0 and 1 which represents the probability that one of the possibilities is present in an image.

3 You are working out error analysis and counting up what errors the algorithm makes. Which of the following do you think you should manually go through and carefully examine, one image at a time?

Answer: 500 images of the dev set, on which the algorithm made a mistake.

Comment: We focus on images that the algorithm got wrong from the dev set. That is the one we use to make choices between different iterations of the system.

4 Version 1: After working on the data for several weeks, your team ends up with the following data:

4 Version 2: 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 $(0, ?1, 1, 1)^Y$ then the learning algorithm will not be able to use that example. True/False?

Answer: False.

Comment: 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 Version 1: 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?

Answer: False.

5 Version 2: 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. How should you split the dataset into train/dev/test sets?

Answer: 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.

Comment: 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 Version 1: Assume you've finally chosen the following split between the data: Training 12%, Training Dev 15.1%, Dev 12.6%, Test 15.8%. You also know that human level error on the road sign and traffic signals classification task is around 0.5%. Which of the following is True?

Answer: You have a high bias.

Comment: The avoidable bias is significantly high since the training error is a lot higher than the human level error.

6 Version 2: Assume you've finally chosen the following split between the data: Training 8.8%, Training Dev 9.1%, Dev 14.3%, Test 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 is True?

Answer: You have a large data mismatch problem because your model does a lot better on the training dev set than on the dev set; You have a large avoidable bias problem because your training error is quite a bit higher than the human level error.

7 Version 1: Assume you've finally chosen the following split between the data: Training 2, Training dev 2.3, Dev, 1.3, Test 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 you conclude that the Bayes error for the dev/test distribution is probably higher than for the train distribution. True/False?

Answer: False

7 Version 2: Based on table from the previous question, a friend thinks that the training data distribution is much easier than the dev/test distribution. What do you think?

Answer: There's insufficient information to tell if your friend is right or wrong.

Comment: The algorithm does better on the distribution of data it trained on. But you don't know if it's because it trained on that no distribution of it is really is easier. To get a better sense, measure human level error separately on both distribution.

8 Version 1: 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, Error due to incorrectly labeled data 4.1, Errors due to foggy pictures 2.0, Error 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?

Answer: 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.

Comment: You should consider the trade off between the data accessibility and potential improvement of your model trained on this additional data.

8 Version 2:The percentages changed to 14.3, 4.1, 8.0, 2.2, 1.0, then The results from this analysis implies that the team's highest priority should be to bring more foggy pictures into the training set so as to address the 8.0% of errors in that category. True/False.

Answer: False because this would depend on how easy it is to add this data and how much you think your team thinks it will help.

9 Version 1: 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, Error due to incorrectly labeled data 4.1, Errors due to foggy pictures 3.0, Error due to partially occluded elements 7.2, Errors due to other causes 1.0. 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?

Answer: The firm will reduce the dev set error with 7.2% at the most.

Comment: 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.

9 Version 2: You can buy a specially designed windshield wiper that help wipe off some of the raindrops 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.

Comment: 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: Which of the following do you agree with? (Check all that apply)

Answer: 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, since human vision is very accurate for the problem you're solving.

Comment: 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 a foggy weather. I will very likely help.

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).

Answer: 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; 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.

Comment: Because you want to make sure that your dev and test data come from the same distribution for your algorithm to make your team's iterative development process efficient; Deep learning algorithms are quite robust to having slightly different train and dev distribution.

12 Version 1: 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. 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?

Answer: False.

Comment: 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.

12 Version 2: 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. Images containing yellow lights are quite rate, and she does not have enough data to build a good model. She hopes you can help her out using transfer learning. What do you tell your colleague?

Answer: She should try using weights pre trained on your dataset, and fine tuning further with the yellow light dataset.

Comment: 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 is after the last hidden layer and initialize it with your trained parameters.

13 Version 1: 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?

Answer: False.

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

13 Version 2: Another colleague wants to use microphones placed outside the car to better hear if there're 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?

Answer: Neither transfer learning nor multi task learning seems promising.

Comment: The problem he is trying to solve is quite different from yours. The different dataset structures make it probably impossible to use transfer learning of multi task learning.

14 Version 1: To recognize red and green lights, you have been using this approach:

- A input an image to a neural network and have it directly learn a mapping to make a prediction as to whether there's a red light and/or green light.

A teammate proposes a different, two step approach:

- B in this two step approach, you would first i detect the traffic light in the image if any, then 2 determine the color of the illuminated lamp in the traffic light.

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

Answer: False.

Comment: A is an end to end approach as it maps directly the input to the output.

14 Version 2: 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 sue an end to end approach?

Answer: There is a large dataset available.

Comment: To get good results when using an end to end approach, it is necessary to have a big dataset.

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

Answer: Large training set.

Comment: In may fields, it has been observed that end to end learning works better in practice, but requires a large amount of data.