← Autonomous driving (case study)

Graded Quiz • 45 min Error Analysis

## Autonomous driving (case study)

Video: Carrying out error analysis 10 min

Video: Cleaning up incorrectly labeled data 13 min

Video: Build your first system quickly, then iterate 6 min

Mismatched training and dev/test set

Video: Training and testing on different distributions 10 min

Video: Bias and Variance with mismatched data distributions
18 min

Video: Addressing data mismatch

Learning from multiple tasks Video: Transfer learning 11 min Video: Multi-task learning 12 min

End-to-end deep learning

Video: What is end-to-end deep learning?
11 min

Video: Whether to use end-to-end deep learning
10 min

Machine Learning flight simulator

Quiz: Autonomous driving (case study) 15 questions

Heroes of Deep Learning (Optional)

Video: Ruslan Salakhutdinov interview

Submit your assignment

DUE DATE Jan 18, 3:59 PM CST ATTEMPTS 3 every 8 hours Try again

Grade View Feedback
100% We keep your highest score 6 P P

Congratulations! You passed! Keep Learning GRADE 100% TO PASS 80% or higher Autonomous driving (case study) LATEST SUBMISSION GRADE To help you practice strategies for machine learning, in 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  $y^{(i)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} \text{"stop sign"}$  "pedestrian crossing sign" "construction ahead sign" "red traffic light" "green traffic light"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 training a basic model and see what mistakes it makes. Spend a few days getting the internet data, so that you understand better what data is available. 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.

False

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.

 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? 10,000 images on which the algorithm made a mistake

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

Correct
Focus on images that the algorithm got wrong. Also, 500 is enough to give you a good initial sense of the error statistics. There's probably no need to look at 10,000, which will take a long time.

 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,  $y^{(i)} = egin{pmatrix} 0 \\ 0 \end{bmatrix}$  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  $\begin{bmatrix} 0\\ ?\\ 1\\ \end{bmatrix}$  then the learning algorithm will not be able to use that example. True/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.

The control of the control of the cost such that it is not influenced by the fact that some entries haven't been labeled.

The control of the cost such that it is not influenced by the fact that it is n

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. How should you split the dataset into rain/del-viets since Mix all the 100,000 images with the 900,000 images you found online. Shuffle everything. Split the 1,000,000 images dataset into 600,000 for the training set, 200,000 for the dev set and 200,000 for the test set. Mix all the 100,000 images with the 900,000 images you found online. Shuffle everything. Split the 1,000,000 images dataset into 980,000 for the training set, 10,000 for the dev set and 10,000 for the test set.

Choose the training set to be the 900,000 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 in dev and test sets. © 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.

Correct
Yes. As seen 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 "real" data to avoid having a data-mismatch problem.

Dataset: Contains:

Pathogen and provided from 1900,000 internet images + 60,000 car's front-facing camera images)

Training
20,000 images randomly picked from 1900,000 internet images + 60,000 car's front-facing camera images)

Training
Dev facing camera images)

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. 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 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 training error is quite higher than the human-level error. 7. 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?

 Your friend is right. (I.e., Bayes error for the training data distribution is probably lower than for the dev/test distribution.) Your friend is wrong. (I.e., Bayes 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 trained on. But you don't know if it's because it trained on that no distribution or if it really is easier. To get a better sense, measure human-level error separately on both distributions.

8. You decide to focus on the deviset and check by hand what are the errors 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 8.0%

Errors due to rain drops stuck on your car's front-facing camera 2.2%

Errors due to other causes 1.0% In this table, 4.1%, 8.0%, etc. are a fraction of the total dev set (not just examples your algorithm mislabeled). For example, about 8.0/15.3 = 52% of your errors are due to foggy pictures. 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?

Additional Note: there are subtle concepts 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 quiz, to understand what issues that you will want to consider when you are building your own machine learning project. False because it depends on how easy it is to add foggy data. If foggy 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.

9. You can buy a specially designed windshield wiper that help wipe off some of the raindrops on the front-facing camera.

1/1 point
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 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.

10. You decide to use data augmentation to address foggy images. You find 1,000 pictures of fog off the internet, and "add" 1/1 point them to clean images to synthesize foggy days, like this: image from foggy image from synthesized front-facing camera the internet foggy image

Which of the following statements do you agree with?

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.

Yes. 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.

After working further on the problem, you've decided to correct the incorrectly labeled data on the deviset. Which of these 1/1 point statements do you agree with? (Check all that apply).

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.

Correct
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 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 quiter are, and she doesn't have enough data to build a good model. She hopes you can help her out using transfer learning.

What do you tell your colleague? You cannot help her because the distribution of data you have is different from hers, and is also lacking the yellow label.

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 is 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 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?

 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. Either transfer learning or multi-task learning could help our colleague get going faster.

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.

Neither transfer learning nor multi-task learning seems promising.

14. To recognize red and green lights, you have been using this approach: (B) In this two-step approach, you would first (i) detect the traffic light in the image (if any), then (ii) determine the color of the illuminated lamp in the traffic light.

15. Approach A (in the question above) tends to be more promising than approach B if you have a \_\_\_\_\_\_ (fill in the blank). 1/1 point

False

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

 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.