Your grade: 100%

Your latest: **100%** • Your highest: **100%**
To pass you need at least 80%. We keep your highest score.

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 an idea 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, this image contains a pedestrian crossing sign and red traffic lights.

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

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, which could be helpful for training even if the distribution of internet data is not the same.

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).

- ○ 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.
- ○ Invest a few days in thinking on potential difficulties, and then some more days brainstorming about possible solutions, before training any model.
- ◉ Train a basic model and do error analysis.
- ○ Spend some time searching the internet for the data most similar to the conditions you expect on production.

Correct
Machine learning projects are most effective when you start with a basic model, analyze its errors, and then iterate to improve it.

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. **1 / 1 point**

Suppose that you use a sigmoid function for the output layer, and the output $\hat{y}$ has shape (5, 1).

Which of the following best describes the cost function?

- ○ $\frac{1}{m}\sum_{i=1}^{m}(-y^{(i)}\log\hat{y}^{(i)}-(1-y^{(i)})\ \log(1-\hat{y}^{(i)})$
- ○ $\frac{\exp \hat{y}_j^{(i)}}{\sum_{j=1}^{5}\exp \hat{y}_j^{(i)}}$
- ◉ $\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{5}\mathrm{L}(\hat{y}_j^{(i)}, y_j^{i})$
- ○ $\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{5}\mathrm{L}(\hat{y}_i^{(j)}, y_i^{(j)})$

Correct

Yes! Here you compare each component (each possible object) of the prediction $\hat{y}$ with the respective component of the label y, and sum over the individual losses. This is appropriate because multiple objects can be present in each image.

3. **True or False:** When trying to determine what strategy to implement to improve the performance of a model, you manually check all images of the training set where the algorithm was successful.

1 / 1 point

- (selected) False
- True

Correct

The training set is typically very large, and manually checking all the successful images would be time-consuming and inefficient. It's more effective to focus on the errors in the dev set to understand where the model needs improvement.

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.

True or False: In multi-task learning, if some examples have missing labels (for example: $\begin{bmatrix} 0 \\ ? \\ 1 \\ 1 \\ ? \end{bmatrix}$), the learning algorithm **cannot** use those examples.

- ( ) True
- (•) False

Correct

As seen in the lecture on multi-task learning, you can compute the cost function in a way that it is not affected by missing labels. The algorithm can still learn from the available labels in the example.

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.

1 / 1 point

How should you split the dataset into train/dev/test sets?

- ( ) 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.
- ( ) 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 80,000 images from your car's front-facing camera. The 20,000 remaining images will be split equally in dev and test sets.
- ( ) 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.

Correct

Yes. 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. 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)	1%
Training-Dev	20,000 images randomly picked from (900,000 internet images + 60,000 car’s front-facing camera images)	5.1%
Dev	20,000 images from your car’s front-facing camera	5.6%
Test	20,000 images from the car’s front-facing camera	6.8%

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?**

- (●) You have a high variance problem.
- ( ) The size of the train-dev set is too large.
- ( ) You have a high bias.
- ( ) You have a large data-mismatch problem.

Correct

The large difference between the training-dev error and the training error indicates that the model is not generalizing well to unseen data from the same distribution, which is a sign of high variance.

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)	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%

Human-level error on this task is approximately 0.5%. (Bayes error is the lowest possible error rate for a task. Human-level error is a good estimation of Bayes error.)

A friend believes the mixed internet/car images (Training) have a lower Bayes error than the car camera images (Dev/Test). What do you think?

- ○ Your friend is likely incorrect.
- ◉ There's insufficient information to determine if your friend is correct or not.

- [ ] Your friend is likely correct.

Correct
To accurately determine and compare Bayes errors, we need to compare the human-level error for each data distribution. The overall human level error does not provide enough information to compare the two distributions.

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?

- [ ] 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.
- [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.
- [ ] 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.

Correct
You should consider the trade-off 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 one of the following statements do you agree with?

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

Correct

You will probably not improve performance by more than 2.2% by solving the raindrops problem.

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 one of the following do you agree with?

- ○ It is irrelevant how the resulting foggy images are perceived by the human eye; the most important thing is that they are correctly synthesized.
- ○ If used, the synthetic data should be added to the training/dev/test sets in equal proportions.
- ○ With this technique, we duplicate the size of the training set by synthesizing a new foggy image for each image in the training set.
- ◉ If used, the synthetic data should be added to the training set.

Correct

The synthetic data can help train the model to get better performance on the dev set, but it 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. Which of the following is a necessary step to take?

1 / 1 point

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

Correct
Recall that the dev set and the test set must come from the same distribution.

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 refer to it as an orange light; however, we will use the US convention of calling it yellow. Images containing yellow lights are quite rare, and she doesn't have enough data to build a good model. She hopes you can help her out using transfer learning.

1 / 1 point

What do you tell your colleague?

- ◉ She should try using weights pre-trained on your dataset and fine-tuning further with the yellow-light dataset.
- ○ If she has (say) 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 “swamping” the yellow lights dataset.
- ○ Recommend that she try multi-task learning instead of transfer learning using all the data.

- [ ] You cannot help her because the distribution of data you have is different from her's and is also lacking the yellow label.

Correct
You have trained your model on a large 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. 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. He has approximately 30,000 examples of each image and 30,000 images without a sign. **1 / 1 point**

True or False: This case could benefit from using multi-task learning.

- [ ] False
- [x] True

Correct
Multi-task learning is suitable here due to the shared high-level features among the required road signs.

14. When building a system to detect cattle crossing a road from images taken with the front-facing camera of a truck, the designers have a large dataset of images. **1 / 1 point**

Which of the following is the strongest reason to use an end-to-end approach?

- [ ] It relies on manually designed components.

- [x] A large dataset is available.
- [ ] It is the default approach for all computer vision tasks.
- [ ] It requires fewer computational resources.

Correct
End-to-end approaches often require substantial datasets to learn complex patterns effectively.

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

First, localize any traffic sign in an image. **After that**, 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**.

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

Correct
When data is limited, a two-step approach can be more effective than training a large end-to-end model.