

Congratulations! You passed!

[Go to next item](#)

Grade received 93.33% Latest Submission Grade 93.33% To pass 80% or higher

- 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 you 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.



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

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

Suppose that you came from working with a project for human detection in city parks, so you know that detecting humans in diverse environments can be a difficult problem. What is the first thing you do? Assume each of the steps below would take about an equal amount of time (a few days).

- Train a basic model and proceed with error analysis.
- Leave aside the pedestrian detection, to move faster and then later solve the pedestrian problem alone.
- Start by solving pedestrian detection, since you already have the experience to do this.
- Spend a few days collecting more data to determine how hard it will be to include more pedestrians in your dataset.

[Expand](#)

Correct

Correct. As discussed in the lecture, it is better to create your first system quickly and then iterate.

- 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. 0 / 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?

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

[Expand](#)



Incorrect

Notice that y_r^s indicates the component r of the example s .

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

1 / 1 point

True

False

Expand



Correct

Correct. This set should be too large to manually check all the images. It is better to focus on the images that the algorithm got wrong from the dev set. Also, choose a large enough subset that we can manually check.

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.

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 $\begin{bmatrix} 0 \\ ? \\ 1 \\ 1 \\ ? \end{bmatrix}$ then the learning

algorithm will not be able to use that example. True/False?

True

False

Expand



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.

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 train/dev/test sets?

1 / 1 point

- 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.
- 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.

- 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.

 [Expand](#)

 **Correct**

Yes. As seen in the 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.

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)	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 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 model is not generalizing well to data from the same training distribution but that it has never seen before.

- 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.

- You have a large data-mismatch problem because your model does a lot better on the training-dev set than on the dev set

 **Correct**

 [Expand](#)

 **Correct**

Great, you got all the right answers.

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)	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 you conclude that the Bayes error for the dev/test distribution is probably higher than for the train distribution. True/False?

True

False

 Expand

 Correct

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 discoveries:

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%

In this table, 4.1%, 8.0%, etc. are a fraction of the total dev set (not just examples of 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.

- First start with the sources of error that are least costly to fix.
- True because it is the largest category of errors. We should always prioritize the largest category of errors as this will make the best use of the team's time.
- True because it is greater than the other error categories added together $8.0 > 4.1 + 2.2 + 1.0$.
- 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.

 Expand

 Correct

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

 Expand

 Correct

Yes. 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:

1 / 1 point



Which of the following do you agree with?

- If used, the synthetic data should be added to the training set.
- With this technique, we duplicate the size of the training set by synthesizing a new foggy image for each image in the training set.
- 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.

 Expand

 Correct

Yes. The synthetic data can help to train the model to get better performance at the dev set, but 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.

1 / 1 point

You have to correct the labels of the test so test and dev sets have the same distribution, but you won't change the labels on the train set because most models are robust enough they don't get severely affected by the difference in distributions. True/False?

- False, the test set should be changed, but also the train set to keep the same distribution between the train, dev, and test sets.
- True, as pointed out, we must keep dev and test with the same distribution. And the labels at training should be fixed only in case

or a systematic error.

- False, the test set shouldn't be changed since we want to know how the model performs in real data.

 Expand

 Correct

Correct! To successfully train a model, the dev set and test set should come from the same distribution. Also, the deep learning models are robust enough to handle a small change in distributions, but if the errors are systematic they can significantly affect the training of the model.

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? 1 / 1 point

- 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.
- Using pre-trained weights can severely hinder the ability of the model to detect dogs since they have too many learned features.
- You should train a single new model for the dogs' task, and leave the previous model as it is.

 Expand

 Correct

Correct. 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? 1 / 1 point

- True
- False

 Expand

 Correct

Correct. 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. 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. We are using multi-task learning. True/False? 1 / 1 point

- False
- True

 Expand



Correct

Correct. Multi-task learning is about joining several tasks that can benefit from each other.

15. An end-to-end approach doesn't require that we hand-design useful features, it only requires a large enough model. True/False?

1 / 1 point

True

False

 **Expand**



Correct

Correct. This is one of the major characteristics of deep learning models, that we don't need to hand-design the features.