

Jalen is training a machine learning model to identify vehicles. He is building a prototype that will distinguish motorcycles from other vehicles.

For his first round of training, Jalen imported two sets of images. He classified the first set of images as “motorcycles.” He classified the second set of images as “non-motorcycles.”

Jalen tested his model using a set of images: **16 classified as motorcycles** and **10 classified as non-motorcycles**. After testing his machine learning model using the 26 training images, he obtained the following experimental data:

- **Motorcycles:**
 - 14 images were correctly identified as motorcycles
 - 2 were incorrectly identified as non-motorcycles.
- **Non-motorcycles:**
 - 6 were correctly identified as non-motorcycles
 - 4 were incorrectly identified as motorcycles.

Using this information, complete the first four columns of Table 1: Training Log - Training 1.

Table 1: Training Log - Training 1

Training Log - Training 1			
Classification	Number Correct	Number Incorrect	Total
Motorcycle			
Non-motorcycle			
Total			

Part 1: Evaluating the Model’s Performance

Jalen is ready to analyze the data from his first round of training.

1. Based on the Training Log results, how likely (unlikely, somewhat likely, very likely, certain) is it that Jalen’s model will correctly identify any image using his categories? Provide evidence to support your answer.

Jalen wants to quantify how well his model performed using **experimental probability**. Experimental probability uses data resulting from an experiment to quantify the likelihood of an event. It is the ratio of the **total number of times an event is observed to the total number of trials in the experiment**.

$$P(\text{Event}) = \frac{n(\text{Event})}{n(\text{Total Trials})}$$

$n(\text{Event})$ is read “n of Event” or “Number of Event” and represents the number of times the event was observed. $n(\text{Total Trials})$ represents the total number of trials in the experiment.

The probability of an event, or $P(\text{Event})$, is a number between 0 and 1. The closer $P(\text{Event})$ is to 1, the more likely it is to happen.

2. Jalen’s event is a correct identification of an image (*Correct*). What is the probability $P(\text{Correct})$ that Jalen’s model will correctly identify an image?

$$P(\text{Correct}) = \frac{n(\text{Correct})}{n(\text{Total Images})} = \underline{\hspace{2cm}}$$

$n(\text{Correct})$ is read “n of Correct” or “Number of Correct” and represents the number of images identified correctly. $n(\text{Total Images})$ represents the total number of images tested.

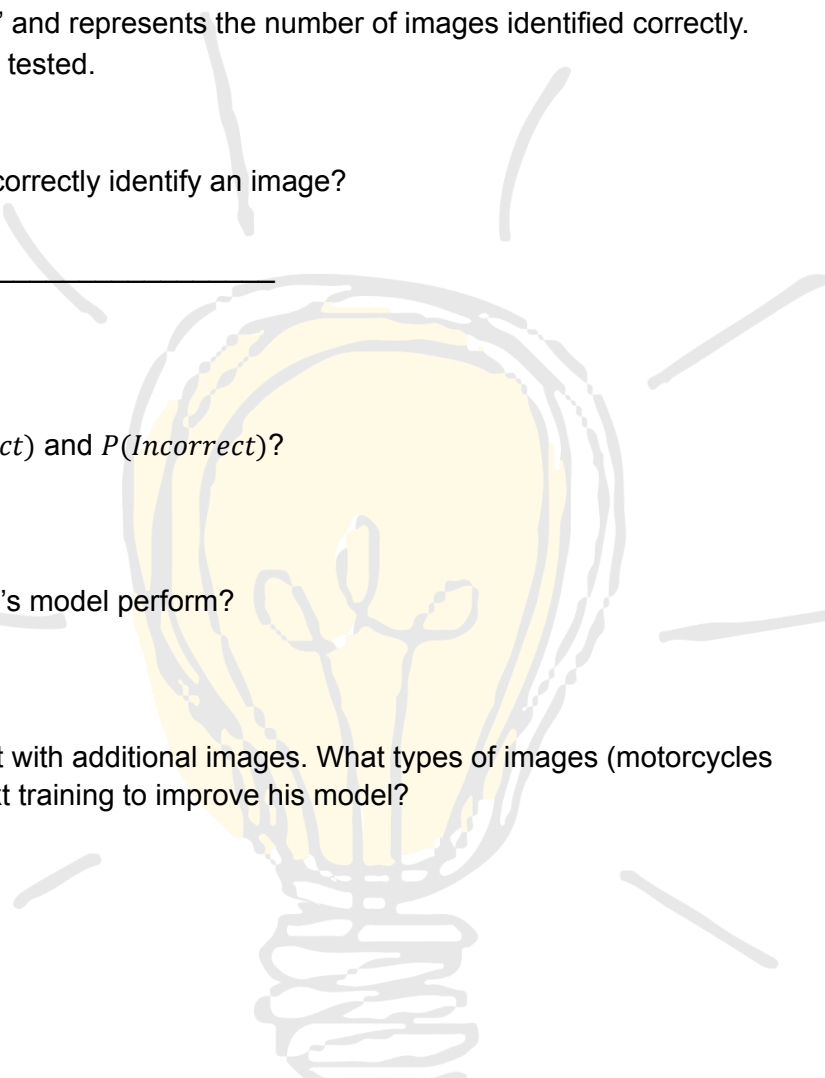
3. What is the probability that Jalen’s model will incorrectly identify an image?

$$P(\text{Incorrect}) = \frac{n(\text{Incorrect})}{n(\text{Total Images})} = \underline{\hspace{2cm}}$$

4. What do you notice about the values of $P(\text{Correct})$ and $P(\text{Incorrect})$?

5. Based on these probabilities, how well did Jalen’s model perform?

6. Jalen wants to improve his model by retraining it with additional images. What types of images (motorcycles or non-motorcycles) should Jalen add in the next training to improve his model?



Part 2: Evaluating the Model's Performance by Type of Image

Jalen wants to know how well his model performed when shown images classified as “non-motorcycle.” This is an example of a **conditional probability**. A conditional probability can be experimental or theoretical. It describes the likelihood of an event when we know that another event has happened.

Jalen wants to know the probability that his model will correctly identify an image, given the image is a non-motorcycle. In other words, when he knows the image is a non-motorcycle. This is written as:

$$P(\text{Correct}|\text{NonMotorcycle})$$

To find this conditional probability, look only at the row titled, “Non-motorcycle.”

1. Find the total number of images classified as non-motorcycle, or $n(\text{NonMotorcycle})$.
2. Find the cell in the table that shows the number of images correctly identified **and** classified as non-motorcycle. This is written as $n(\text{Correct} \cap \text{NonMotorcycle})$. The symbol \cap means that you need to meet multiple criteria; in this case, it tells you that any image included must be both correctly identified and be classified as “non-motorcycle.”
3. The conditional probability is the ratio of the number of images included in the set $\text{Correct} \cap \text{NonMotorcycle}$ to the number of images classified as non-motorcycle. Or,

$$P(\text{Correct}|\text{NonMotorcycle}) = \frac{n(\text{Correct} \cap \text{NonMotorcycle})}{n(\text{NonMotorcycle})}$$

$$P(\text{Correct}|\text{NonMotorcycle}) = \underline{\hspace{4cm}}$$

4. What does $P(\text{Correct}|\text{NonMotorcycle})$ tell you about Jalen's machine learning model? What types of images should Jalen add to his training set to improve the model?

Part 3: Evaluating the Model's Performance by Subcategories

Jalen is reviewing the images he used to test his model. He groups these images using the following subcategories:

Motorcycle:

- Standard Motorcycle: 12 tested, 12 correctly identified
- Sport Motorcycle: 4 tested, 2 correctly identified

Non-motorcycle:

- Truck: 4 tested, 4 correctly identified
- Bicycle: 4 tested, 2 correctly identified
- Scooter: 2 tested, 0 correctly identified

Complete Table 2 using these data.

Table 2: Results of Training 1 with Subcategories

Results of Training 1 with Subcategories			
Category	Number Correct	Number Incorrect	Total
Motorcycle			
Standard			
Sport			
Non-motorcycle			
Truck			
Bicycle			
Scooter			
Total			

1. What is the probability of the model correctly identifying an image, given the image is of a truck?

$$P(\text{Correct}|\text{Truck}) =$$

2. What is the probability of the model correctly identifying an image, given the image is of a sport motorcycle?
Write an equation using the correct notation for conditional probability.

3. What is the probability of the model correctly identifying an image, given the image is of a scooter? Write an equation using the correct notation for conditional probability.
4. What types of images should Jalen add to his training set to improve the model? Why?
5. How can conditional probability help you improve your machine learning model?

Now that you understand conditional probabilities, you can go calculate your model's conditional probabilities in the Machine Learning Training Log!

