

# Probability II

Michael C. Hackett

Assistant Professor, Computer Science

Community  
College  
*of* Philadelphia

# Marginal Probability

- The contingency table below is for a hypothetical test of a machine learning algorithm's ability to predict if an image is a picture of a car.
- We could use the contingency table to find probabilities.

		Image		Total
		Car	Not a car	
ML Algorithm	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

# Marginal Probability

- Probability the ML algorithm classified a picture as being an image of a car:

$$P(\text{Prediction is Car}) = \frac{195}{851} \sim 0.23$$

		Image		Total
		Car	Not a car	
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

# Marginal Probability

- Probability of a picture being an image of a car:

$$P(\text{Image is Car}) = \frac{310}{851} \sim 0.36$$

		Image		Total
		Car	Not a car	
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

# Marginal Probability

- Both are examples of a **marginal probability**.
  - Marginal probabilities are based only on one variable

$P(\text{Prediction is Car})$

- Probability is based only on the Algorithm's choice

$P(\text{Image is Car})$

- Probability is based only on the content of the Image

# Joint Probability

- Probability the ML algorithm classified a picture of a car and the Image was of a car:

$$P(\text{Prediction is Car and Image is Car}) = \frac{178}{851} \sim 0.21$$

		Image		Total
		Car	Not a car	
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
Total		310	541	851

- This is an example of a **joint probability**.
  - Probabilities based on two or more variables

# Joint Probability

- Probability the ML algorithm classified a picture as not of a car and the Image was not of a car:

$$P(\text{Prediction is Not a car, Image is Not a car}) = \frac{524}{851} \sim 0.62$$

		Image		Total
		Car	Not a car	
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

# Joint Probability

- The contingency table converted to proportions

		Image		Total
		Car	Not a car	
ML Algorithm Prediction	Car	0.21	0.02	0.23
	Not a car	0.15	0.62	0.77
	Total	0.36	0.64	1.00



# Joint Probability

- Joint probability distribution

Joint outcome	Probability
Prediction is Car, Image is Car	0.21
Prediction is Car, Image is Not a car	0.02
Prediction is Not a car, Image is Car	0.15
Prediction is Not a car, Image is Not a car	0.62
Total	1.00

- Does not indicate the accuracy of the algorithm

# Conditional Probability

- Of the images that were actually an image of a car, what was the probability that ML algorithm classified it as a car?:

$$P(\text{Prediction is Car given Image is Car}) = \frac{178}{310} \sim 0.57$$

		Image		
		Car	Not a car	Total
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

- This is an example of a **conditional probability**.
  - One probability based on another probability.
  - “Given” is often written “|”

# Conditional Probability

- The conditional probability of A given B is computed using:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

- $P(\text{Prediction is Car} \mid \text{Image is Car})$
- $P(A \text{ and } B) = 0.2092$  (from Slide 6)
- $P(B) = \text{Image is Car} = 0.3643$  (from Slide 4)

$$P(\text{Prediction is Car} \mid \text{Image is Car}) = \frac{0.2092}{0.3643} \sim 0.57$$

- Same result from the previous slide

# Conditional Probability

- Of the images the ML algorithm classified as a car, how many were actually an image of a car?:

$$P(\text{Image is Car} \mid \text{Prediction is Car}) = \frac{178}{195} \sim 0.91$$

		Image		Total
		Car	Not a car	
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

# Conditional Probability

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

- $P(\text{Image is Car} \mid \text{Prediction is Car})$
- $P(A \text{ and } B) = 0.2092$  (from Slide 8)
- $P(B) = \text{Prediction is Car} = 0.23$  (from Slide 3)

$$P(\text{Image is Car} \mid \text{Prediction is Car}) = \frac{0.21}{0.23} \sim 0.91$$

- Same result from the previous slide

# Conditional Probability

$$P(\text{Prediction is Car} \mid \text{Image is Car}) = \frac{0.2092}{0.3643} \sim 0.57$$

The probability the ML algorithm will predict “Car” given that the image is a car: ~57%

- In other words: If an image is of a Car, there is a ~57% probability the algorithm will correctly predict Car

$$P(\text{Image is Car} \mid \text{Prediction is Car}) = \frac{0.21}{0.23} \sim 0.91$$

The probability that an image is of a car given that the ML algorithm predicted “Car”: ~91%

- In other words: If the algorithm predicted Car, there is a ~91% probability it was correct (the image is of a Car)

# General Multiplication Rule

- Recall from the previous lecture that the multiplication rule is used for calculating probabilities for multiple independent events occurring:

- Where A and B are two different and independent events

$$P(A \text{ and } B) = P(A) \times P(B)$$

- The **General Multiplication Rule** allows calculating probabilities for events that might not be independent:

- Where A and B are two different events

$$P(A \text{ and } B) = P(A | B) \times P(B)$$

# General Multiplication Rule

- We previously determined that:

$$P(\text{Prediction is Car and Image is Car}) = \frac{178}{851} \sim 0.21$$

		Image		
		Car	Not a car	Total
ML Algorithm Prediction	Car	178	17	195
	Not a car	132	524	656
	Total	310	541	851

- We can calculate/verify this using only the General Multiplication Rule



# General Multiplication Rule

$$P(A \text{ and } B) = P(A | B) \times P(B)$$

- $P(\text{Prediction is Car} | \text{Image is Car})$
- $P(A | B) = 0.57$  (from Slide 11)
- $P(B) = \text{Image is Car} = 0.3643$  (from Slide 4)

$$P(\text{Prediction is Car and Image is Car}) = 0.57 \times 0.364 \sim 0.21$$

- Same result from the previous slide

# General Multiplication Rule

- The formula previously shown for calculating conditional probabilities is simply a re-arrangement of the general multiplication rule.

- General Multiplication Rule

$$P(A \text{ and } B) = P(A | B) \times P(B)$$

- Conditional Probability Formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

# The Gambler's Fallacy

- In the previous lecture, we saw that the events of a Roulette wheel are independent.
  - Knowing the previous number has no effect on the next number.
- Let's say 8 red numbers came up in a row.
  - The probability of 8 red numbers showing up in a row is
$$\left(\frac{18}{37}\right)^8 \sim 0.003 \sim 0.3\%$$
- The probability of 9 red numbers showing up in a row is
$$\left(\frac{18}{37}\right)^9 \sim 0.001 \sim 0.1\%$$
- With only a 0.1% probability of 9 red numbers appearing in a row, one might think there is an almost definite chance the next number will be black

# The Gambler's Fallacy

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$P(\textit{Black} | 8 \textit{ Reds in a row}) = \frac{P(\textit{Black and 8 Reds in a row})}{P(8 \textit{ Reds in a row})}$$

$$P(8 \textit{ Reds in a row}) = \left(\frac{18}{37}\right)^8 \sim 0.003137$$

$$P(\textit{Black}) = \frac{18}{37} \sim 0.486486$$

$$P(\textit{Black and 8 Reds in a row}) = P(\textit{Black}) \times P(8 \textit{ Reds in a row}) \sim 0.001526$$

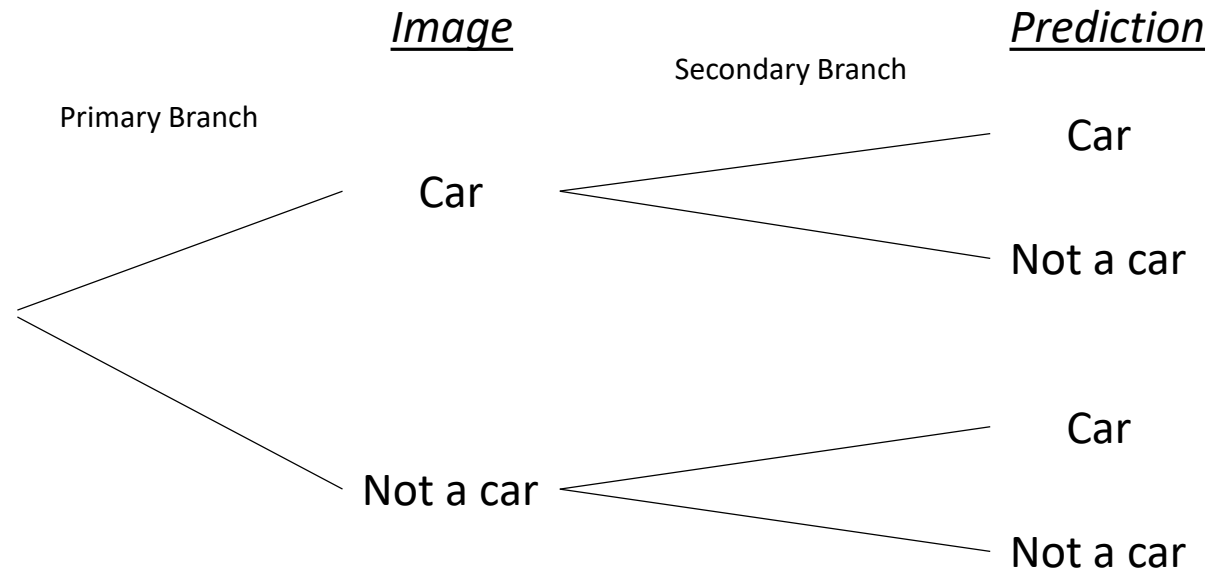
$$P(\textit{Black} | 8 \textit{ Reds in a row}) = \frac{0.001526}{0.003127} \sim \mathbf{.486}$$

# The Gambler's Fallacy

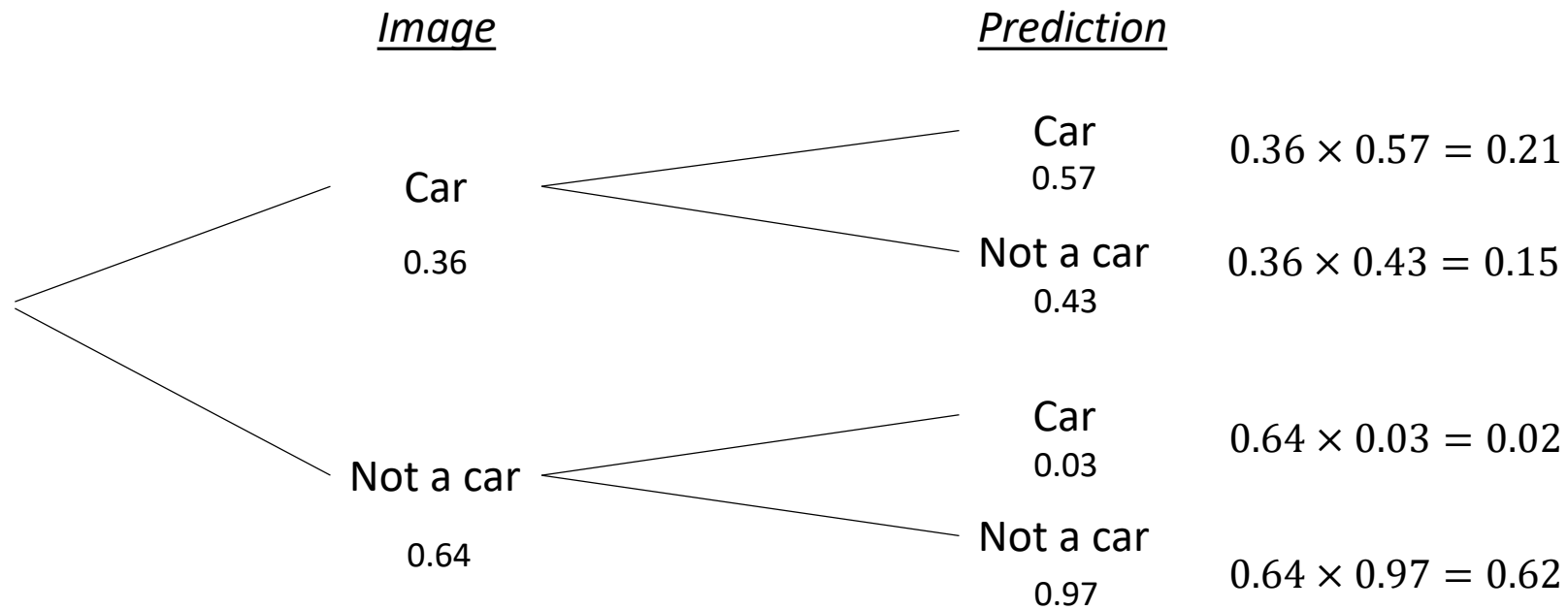
- The probability of a black number given 8 reds appearing in a row is still .486
  - The same probability we determined for a black number appearing on any spin
- Thus, whether no preceding spins, 8 red numbers preceding a spin, or 100 red numbers preceding a spin, the probability of a black number appearing on the next spin is still .486
  - Mathematically shows that each event resulting from a spin is independent.

# Tree Diagrams

- Tree diagrams visually organize outcomes and probabilities.



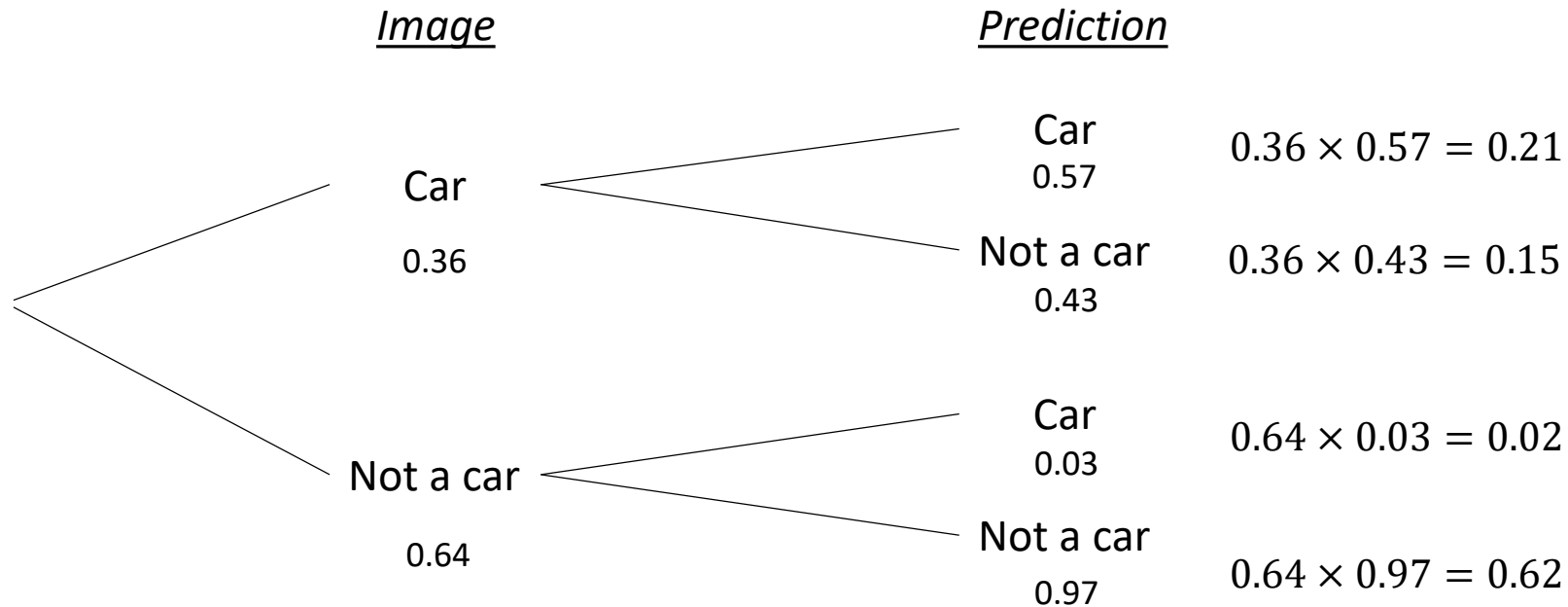
# Tree Diagrams



# Tree Diagrams

$$P(\text{Prediction is Car} \mid \text{Image is Car}) = 0.57$$

$$P(\text{Image is Car and Prediction is Car}) = 0.21$$

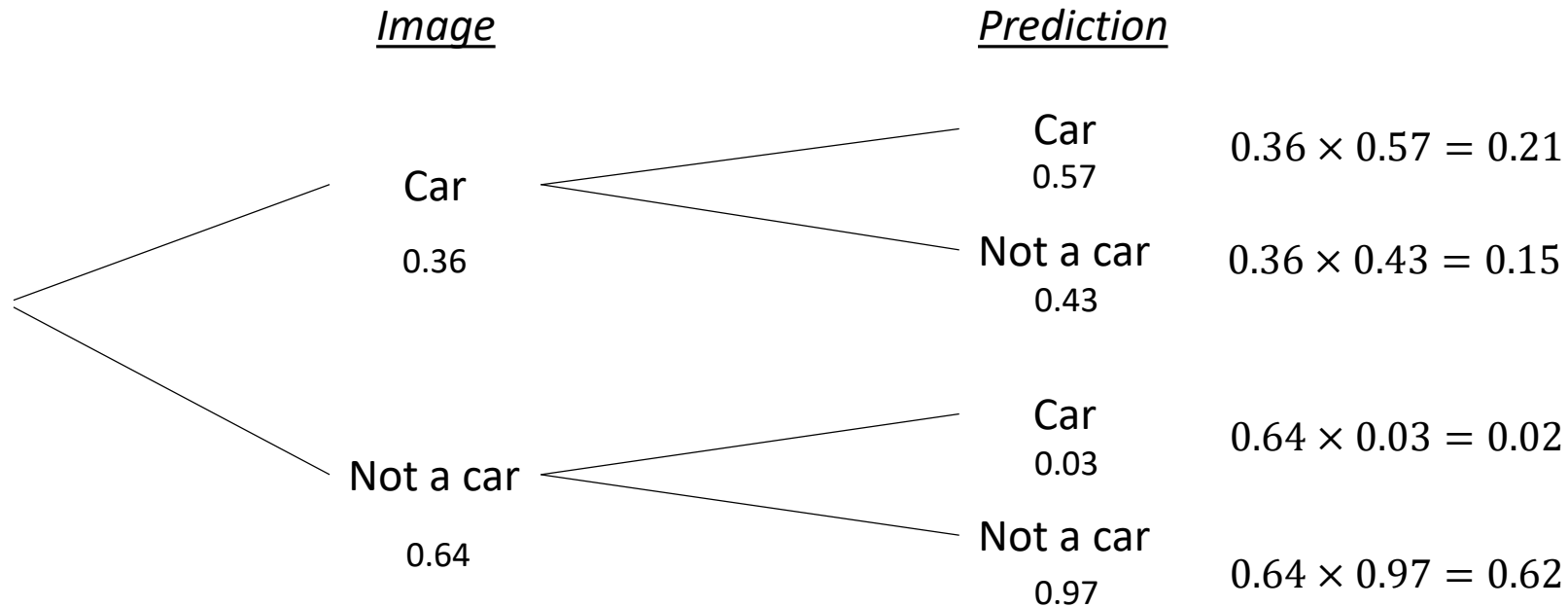




# Tree Diagrams

$$P(\text{Prediction is Car} \mid \text{Image is Not a car}) = 0.03$$

$$P(\text{Image is Not a car and Prediction is Car}) = 0.62$$



# Tree Diagrams

- The last slides gave the conditional probability

$$P(\text{Prediction is Car} \mid \text{Image is a car})$$
$$P(\text{Prediction is Car} \mid \text{Image is Not a car})$$

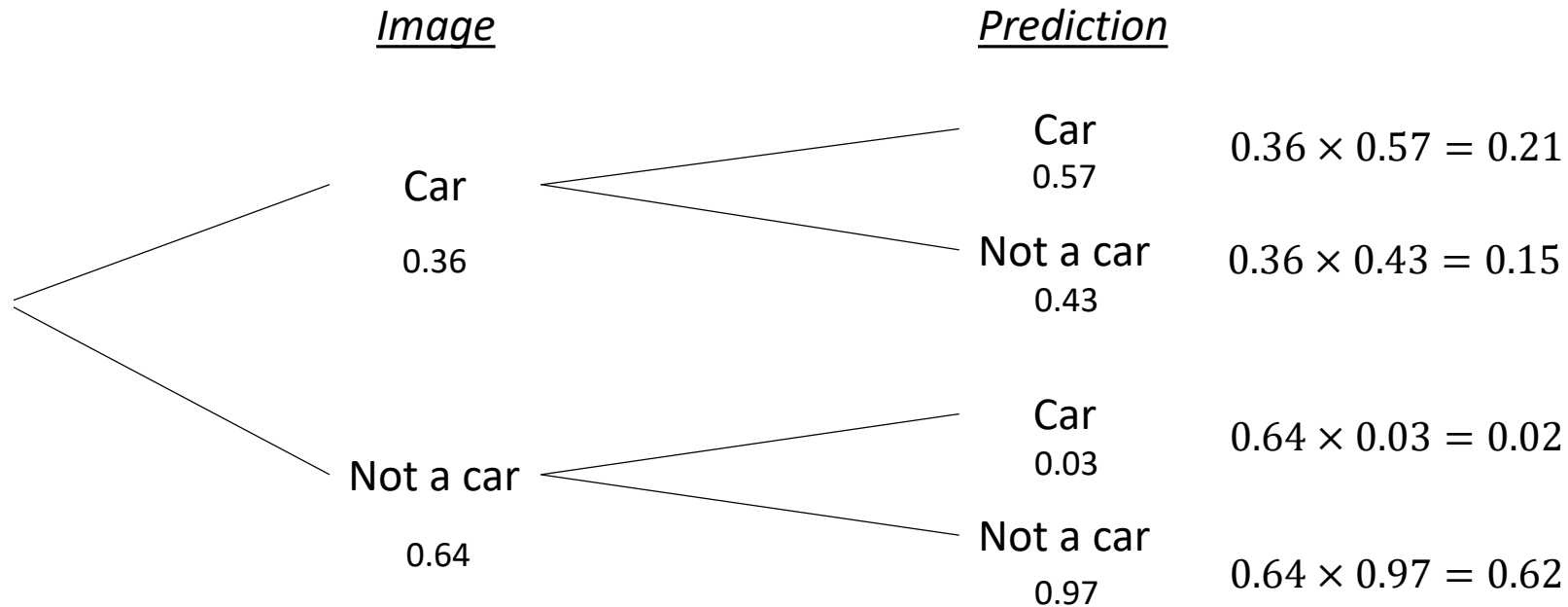
- What if we wanted to find the inverse?

$$P(\text{Image is a car} \mid \text{Prediction is Car})$$
$$P(\text{Image is Not a car} \mid \text{Prediction is Car})$$

# Tree Diagrams

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$P(\text{Image is a car} \mid \text{Prediction is a car}) = \frac{0.21}{0.23} = 0.91$$



# Bayes' Theorem

- Bayes' Theorem is a generalization of conditional probability

$$P(A_1 | B) = P(\text{Image is a car} | \text{Prediction is a car})$$

$$P(A_1 | B) = \frac{P(B | A_1) \times P(A_1)}{P(B | A_1) \times P(A_1) + P(B | A_2) \times P(A_2) + \cdots + P(B | A_n) \times P(A_n)}$$

- Where  $A_1, A_2, \dots, A_n$  are the different outcomes of the first variable

$$P(A_1 | B) = P(\text{Image is a car} | \text{Prediction is a car})$$

# Bayes' Theorem

- To apply Bayes' Theorem:
  - First, identify the marginal probabilities of each possible outcome of the first variable:  $P(A_1), P(A_2), \dots, P(A_n)$ 
    - $P(A_1) = P(\text{Image is a car}) = 0.36$
    - $P(A_2) = P(\text{Image is Not a car}) = 0.64$

$$P(A_1 | B) = P(\text{Image is a car} | \text{Prediction is a car})$$

# Bayes' Theorem

- To apply Bayes' Theorem:
  - Second, identify the probability of the outcome B, conditioned on each possible scenario for the first variable:  $P(B | A_1), P(B | A_2), \dots, P(B | A_n)$

$$P(B | A_1) = P(\text{Prediction is a car} | \text{Image is a car}) = \frac{P(\text{Prediction is a car and Image is a car})}{P(\text{Image is a car})} = 0.57$$

$$P(B | A_2) = P(\text{Prediction is a car} | \text{Image is Not a car}) = \frac{P(\text{Prediction is a car and Image is Not a car})}{P(\text{Image is a car})} = 0.03$$

# Bayes' Theorem

$$P(A_1 | B) = P(\text{Image is a car} | \text{Prediction is a car})$$

$$P(A_2 | B) = P(\text{Image is Not a car} | \text{Prediction is a car})$$

$$P(B | A_1) = P(\text{Prediction is a car} | \text{Image is a car})$$

$$P(B | A_2) = P(\text{Prediction is a car} | \text{Image is Not a car})$$

$$P(B | A_1) = \frac{P(B | A_1) \times P(A_1)}{P(B | A_1) \times P(A_1) + P(B | A_2) \times P(A_2) + \dots + P(B | A_n) \times P(A_n)}$$

$$P(A_1 | B) = \frac{\mathbf{0.57 \times 0.36}}{0.57 \times 0.36 + 0.03 \times 0.64} = \frac{0.21}{0.21 + 0.02} = \frac{0.21}{0.23} = 0.91$$

# Bayes' Theorem

$$P(A | B) = P(\text{Image is a car} | \text{Prediction is a car})$$

$$P(A) = P(\text{Image is a car})$$

$$P(B) = P(\text{Prediction is a car})$$

- Bayes' Theorem can be simplified to:

$$P(A | B) = \frac{P(B | A) \times P(A)}{P(B)}$$

$$P(A | B) = \frac{0.57 \times 0.36}{0.23} = \frac{0.21}{0.23} = 0.89$$