



# **CIS 635 - Knowledge Discovery & Data Mining**

Basics of Probability



## Standard vs AI solutions

- Find the maximum from a given list of numbers.
- Whats was the recorded lowest temperature of Grand Rapids in last 20 years?



## Standard vs AI solutions

- Find the maximum from a given list of numbers.
- Whats was the recorded lowest temperature of Grand Rapids in last 20 years?
- Will it snow tomorrow? What are the chances?
- Who is going to win in the next NFL games between Texans vs Ravens?
- Translate the following into “French”:
  - *“English and French are two European languages; they have a lot in common; however they also possess a lot of differences, especially when we talk about conjugations, contractions and gender usages.”*



# Standard vs AI solutions

- Difference between a regular/standard and an AI/ML algorithm
  - Regular/standard algorithms (Deterministic)
  - AI/ML algorithms (non deterministic, lives with uncertainty)
  - Probability theory is the branch of Math that talks about uncertainty.



## Standard vs AI solutions

- Fill in the Gaps
  - I am very .....???

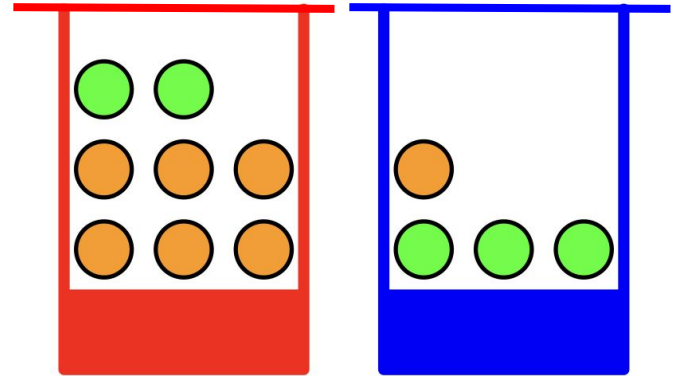


## Standard vs AI solutions

- Fill in the Gaps
  - It's already 7pm; I have been working since early morning, I am very .....??? .....

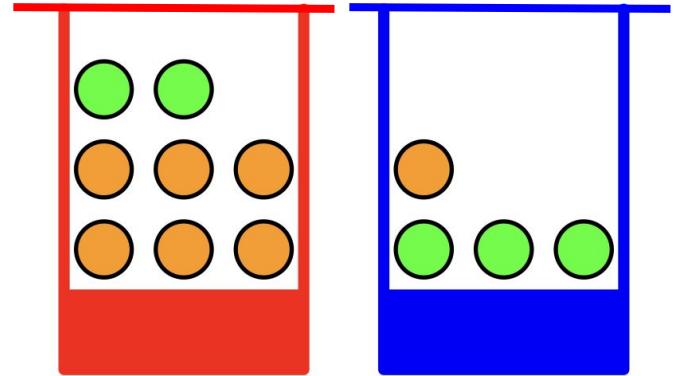
# Basics of Probability

- Let's say, we have two boxes of mixed oranges and apples (as depicted in the right)



# Basics of Probability

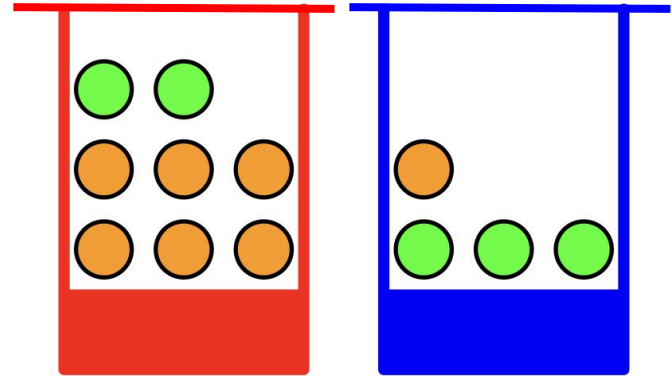
- Let's say, we have two boxes of mixed oranges and apples (as depicted in the right)
- If you are asked to chose a Box (before you see what's inside) which one will you choose?





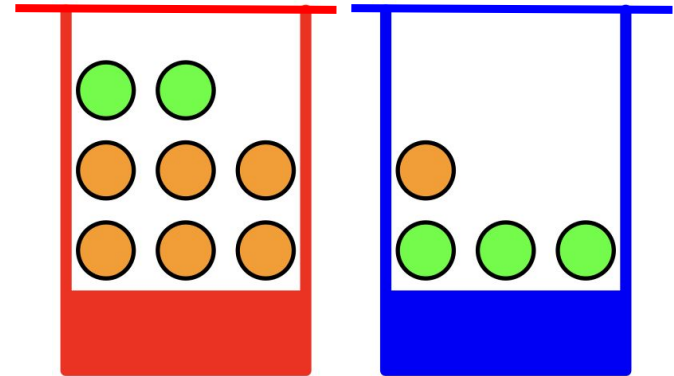
# Basics of Probability

- Let's say, we have two boxes of mixed oranges and apples (as depicted in the right)
- If you are asked to choose a Box (before you see what's inside) which one will you choose?
- If you don't have any **Bias** (say color, or the location of the boxes, etc), assigning a 50%-50% preference is a reasonable assumption



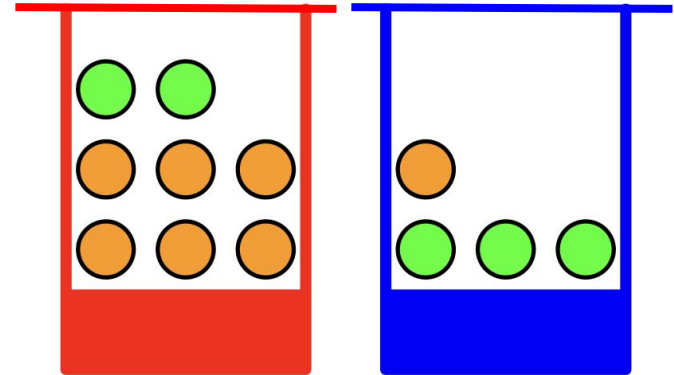
# Basics of Probability

- Let's say, we have two boxes of mixed oranges and apples (as depicted in the right)
- If you are asked to choose a Box (before you see what's inside) which one will you choose?
- If you don't have any bias (say color, or the location of the boxes, etc), assigning a 50%-50% preference is a reasonable assumption
- This is generally known as “**Uniform distribution**” assumption



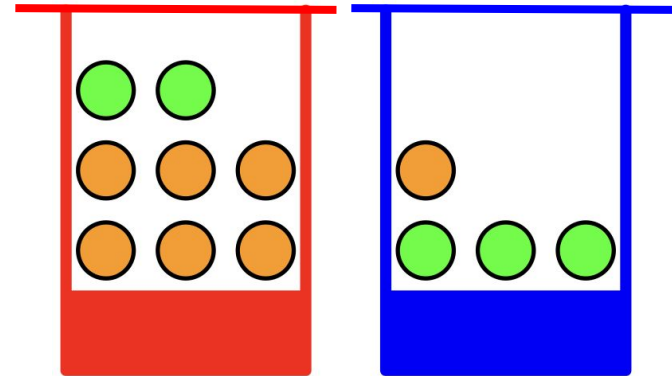
# Basics of Probability

- Let's say, we have two boxes of mixed oranges and apples (as depicted in the right)
- For some unknown reason when some people were asked to choose a fruit from one these two (2) boxes, people preferred the **Blue box (60%)** and the **Red one (40%)** of the time.



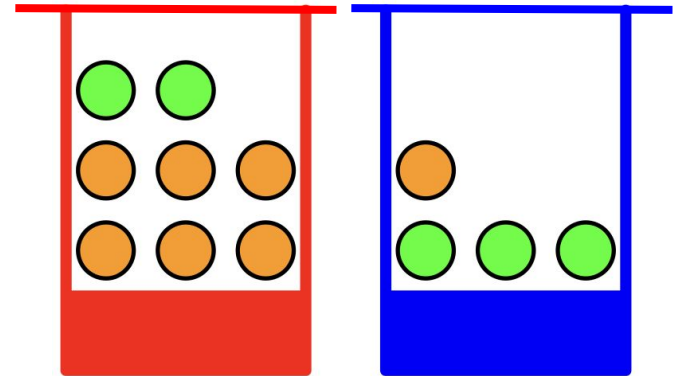
# Basics of Probability

- For some unknown reason when some people were asked to choose a fruit from one these two (2) boxes, people preferred the Blue box (60%) and the red one (40%) of the time.
  - So we have Two **Random variables**:
    - Box (B), and
    - Fruit (F)
  - Probability of choosing the Red box is,  
 $p(B=r) = 40/100 = 0.4$
  - Probability of choosing the Blue box is,  
 $p(B=b) = 60/100 = 0.6$



# Basics of Probability

- For some unknown reason when some people were asked to choose a fruit from one these two (2) boxes, people preferred the Blue box (60%) and the red one (40%) of the time.
  - So we have Two **Random variables**:
    - Box (B), and
    - Fruit (F)
  - Probability of choosing the Red box is,  $p(B=r) = 40/100 = 0.4$
  - Probability of choosing the Blue box is,  $p(B=b) = 60/100 = 0.6$
  - $p(B=r) + p(B=b) = 1$  (**Summation rule**)
    - *Two boxes are completely independent and disjoint*

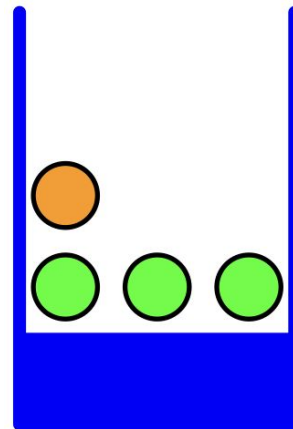


# Basics of Probability

Let's open the boxes, and explore the fruits inside, and explain the content in terms of probabilities.

- $p(F=a|B=b) = \frac{3}{4} = 0.75$
- $p(F=o|B=b) = \frac{1}{4} = 0.25$
- $p(F=a|B=b) + p(F=o|B=b) = 1$

Uniform distribution assumption

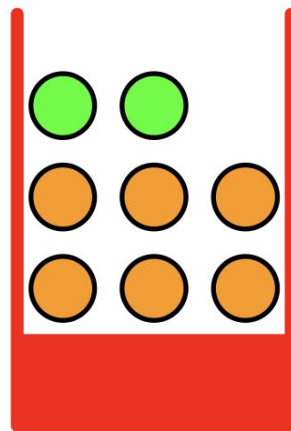


# Basics of Probability

Let's open the boxes, and explore the fruits inside, and explain the content in terms of probabilities.

- $p(F=a|B=r) = 2/8 = 0.25$
- $p(F=o|B=b) = 6/8 = 0.75$
- $p(F=a|B=r) + p(F=o|B=r) = 1$

Uniform distribution assumption

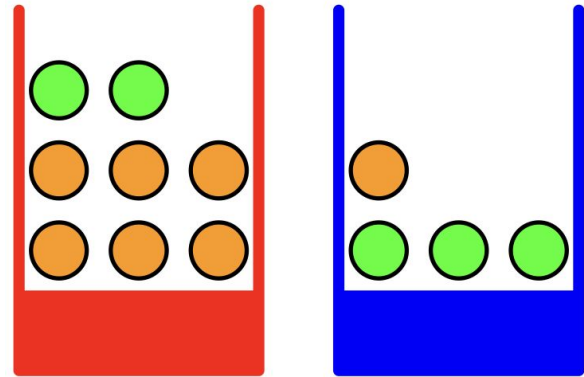


# Basics of Probability

Let's open the boxes, and explore the fruits inside, and explain the content in terms of probabilities.

$$\begin{aligned} p(F=a) &= p(F=a|B=r) p(B=r) + p(F=a|B=b)p(B=b) \\ &= 0.25*0.4 + 0.75*0.6 \\ &= 0.55 \end{aligned}$$

Uniform distribution assumption



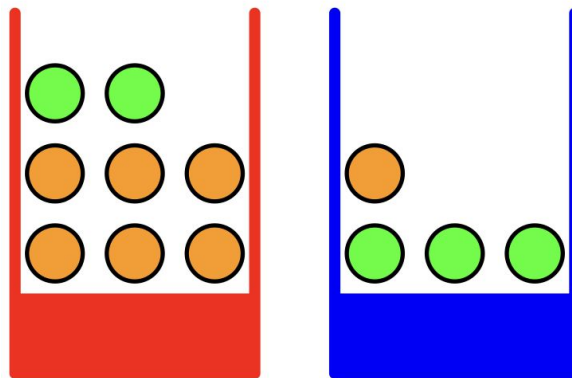


# Basics of Probability

$$p(B = r|F = o) = \frac{p(F = o|B = r)p(B = r)}{p(F = o)} :$$

$$= \frac{3}{4} \times \frac{4}{10} \times \frac{20}{9} = \frac{2}{3}.$$

Bayes' rule/theorem





# What we have learned today

- Differences between Standard vs AI solutions
- Uncertainty is attached to AI solutions
- Probability theory is to a tool to explain and manipulate Uncertainty
- Some Basics of Probability theory
  - Summation Rule
  - Bayes Rule
- Our Probability journey starts here