# Fundamentals for Machine Learning

Syed Arham Naqvi

July 2, 2024

## Contents

# 1 Introduction

What follows is an overview of the fundamental concepts from probability and statistics required to develop a strong understanding of machine learning. I will attempt to summarize and articulate my understanding of the materials/examples presented in the Stanford CS109 course readings at **https://chrispiech.github.io/probabilityForComputerScientists**. Ideally my work here can serve as a solid reference for future studies in Data Science.

# 2 Core Probability

## 2.1 Counting

**Definition 2.1.1 Step Rule of Counting (aka Product Rule of Counting):** If an experiment has two parts, where the first part can result in one of $m$ outcomes and the second part can result in one of $n$ outcomes regardless of the outcome of the first part, the total number of outcomes is $m \cdot n$.

So if the outcome of the first part is from set $A$ where $|A| = m$ and the outcome from the second part is from set $B$ where $|B| = n$, then given that the first outcome in no way influences the second outcome, there must be $m \cdot n$ total outcomes.

**Example 2.1.2:** Assuming the true color model in which each pixel can be $2^{24} \approx 17$million colours, how many distinct pictures can be generated by a) a smartphone camera with 12 million pixels, b) a grid with 300 pixels and c) a grid with 12 pixels?

**Solution 2.1.2:** If each generated image is an experiement, then each pixel would be a single *part* or *step* of the experiment. Since the color of one pixel does not influence that of another, each step is independent. Since every pixel can be one of $2^{24} \approx 17$million color outcomes, the total number of generated images (experiment outcomes) must be $(17\text{million})^n$ where $n$ is the number of pixels.

a. 12 million pixels means $n = 12000000$ so there are $(17\text{million})^{12000000} \approx 10^{86696638}$

b. 300 pixels means $n = 300$ so there are $(17\text{million})^{300} \approx 10^{2167}$

c. 12 pixels means $n = 12$ so there are $(17\text{million})^{12} \approx 10^{86}$

■

**Definition 2.1.3 Mutually Exclusive Counting:** If the outcome of an experiment can either be drawn from set $A$ **or** set $B$ where $|A \cap B| = 0$ (mutual exclusion), there are $|A \cup B| = |A| + |B|$ outcomes in the experiment.

**Example 2.1.4:** A route finding algorithm needs to find routes from Nairobi to Dar Es Salaam. It finds routes that either pass through Mt Kilimanjaro or Mombasa. There are 20 routes that pass through Mt Kilimanjaro, 15 routes that pass through Mombasa and 0 routes passing through both Mt Kilimanjaro and Mombasa. How many routes are there total?

**Solution 2.1.4:** Let $A$ be the set of routes through Mt. Kilimanjaro where $|A| = 20$ and let $B$ be the set of routes through Mombasa where $|B| = 15$. Since there are no routes that pass through both, we know $|A \cap B| = 0$. So as per mutually exclusive counting,

$$\begin{aligned}\text{total outcomes (routes)} &= |A \cup B| \\ &= |A| + |B| \\ &= 20 + 15 \\ &= 35.\end{aligned}$$

■

**Definition 2.1.5 Inclusive Exclusion Counting (aka Sum Rule of Counting):** If the outcome of an experiment can either be drawn from set $A$ **or** set $B$ where $|A \cap B| \neq 0$ (intersection exists), there are $|A \cup B| = |A| + |B| - |A \cap B|$ outcomes in the experiment.

**Example 2.1.6:** An 8-bit string (one byte) is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

**Solution 2.1.6:** Let $A$ be the set of strings beginning with "01". Since the first two bits of the 8-byte string are fixed, only the remaining 6 bits can vary, and so it must be that $|A| = 2^6 = 64$. A similar argument can be made for set $B$; consisting of all strings ending in "10". For both sets however, we have double counted strings beginning with "01" **and** ending with "10". There must be $2^4 = 16$ such strings as 4 of the 8 bits are fixed and so $|A \cap B| = 16$. So as per Sum Rule of Counting,

$$\begin{aligned}
\text{total strings} &= |A \cup B| \\
&= |A| + |B| - |A \cap B| \\
&= 64 + 64 - 16 \\
&= 112
\end{aligned}$$

■

Note that definition 2.1.3 is just a special case of definition 2.1.5 when $|A \cap B| = 0$.

## Overcounting and Correcting

For more difficult counting problems that introduce one or more constraints on the universal set of possibilities, one strategy is to overcount first and then subtract the amount by which we have overcounted. Take for example a 4x4 grid of pixels where each pixel can be either white or blue. If we wanted to count all the configurations with an odd number of blue pixels and where horizontal mirrors are considered indistinct, it would be challenging to pose the problem as a sum of mutually exclusive counts. We can instead start by counting all possible configurations which is $2^4 = 16$. We then note that half of these grids will contain an even number of blue pixels and half will contain an odd number leaving us with $\frac{2^4}{2} = 8$. Finally, we notice that each of the remaining 8 configurations with an odd number of blue pixels have a horizontally-flipped counterpart meaning the number of distinct configurations based on our criteria is actually $\frac{2^4}{2 \cdot 2} = 4$ which is the correct answer.

## 2.2 Combinatorics

**Definition 2.2.1 Permutation Rule:** A permutation is an ordered arrangement of $n$ distinct objects. Those objects can be permuted $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \ldots 2 \cdot 1 = n!$ ways.

**Example 2.2.2:** How many unique orderings of characters are possible for the string "BAYES"?

**Solution 2.2.2:** Since all 5 objects or letters are distinct in this case, the total ways to permute "BAYES" where order is important would be $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ ■

**Definition 2.2.3 Permutations of In-Distinct Objects:** Generally, when there are $n$ objects and:

$$
\begin{aligned}
&n_1 \text{ are the same and} \\
&n_2 \text{ are the same and} \\
&\ldots \\
&n_r \text{ are the same,}
\end{aligned}
$$

then the number of distinct permutations is,

$$\frac{n!}{n_1! \cdot n_1! \cdot \cdots \cdot n_r!}$$

.