Fundamentals of Machine Learning

Syed Arham Naqvi

July 3, 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 of the grid 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 resulting in a double-count. This means that the number of distinct configurations based on our criteria is actually $\frac{2^4}{2 \cdot 2} = 4$.

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:

$$n_1 \text{ are the same and}$$
$$n_2 \text{ are the same and}$$
$$\ldots$$
$$n_r \text{ are the same,}$$

then the number of distinct permutations is:

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

Example 2.2.4:   How many *distinct* orderings of characters are possible for the string "MISSISSIPPI"?

Solution 2.2.4:   If each character was unique we would have 11! distinct orderings. Since therer are 4 S's, 4 I's and 2 P's, we use the permutations of In-Distinct Objects formula to get the number of distinct orderings:

$$\frac{11!}{2! \cdot 4! \cdot 4! \cdot 1!} = 34650$$

∎

DEFINITION 2.2.5 Combinations: A combination is an unordered selection of r objects from a set of n objects. If all objects are distinct and are removed from the pool of potential choices once chosen, then the number of ways of making the overall unordered selection is:

$$\frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Each unique selection of r items must have a corresponding selection of $(n-r)$ ignored items. Since r items have $r!$ permutations and $(n-r)$ items have $(n-r)!$ permutations, the value $n!$ is over-counting our unique selection of r by a factor of $r!(n-r)!$. This is because every one of the $r!$ permutations is being counted $(n-r)!$ times for the number of ways the ignored items can be permuted.

Example 2.2.6:   How many ways are there to select 3 books if there are two books that should not both be chosen together? For example, if you are choosing 3 out of 6 probability books, don't choose both the 8th and 9th edition of the Ross textbook.

Solution 2.2.6:   There are two approaches:

1. we split the problem into three cases and then use the step rule of counting to count the disjoint sets:

   a) choose 2 books from a set of 4 having already chosen the 8th edition and excluding the 9th edition from the pool giving $\binom{5}{2}$

   b) chose 2 books from a set of 4 having already chosen the 9th edition and excluding the 8th edition from the pool giving $\binom{5}{2}$

   c) choosing 3 books from a set of 4 excluding both 8th and 9th editions from the pool giving $\binom{4}{3}$

   $\therefore$ Total $= 2 \cdot \binom{5}{2} + \binom{4}{3} = 16$.

2. We count the total number of ways to choose 3 books from 6 with no constraints. We then count the number of ways our choices could contain both 8th and 9th editions and subtract this value from the total. To count the number of selections to subtract, assume the 8th and 9th addition have already been chosen and vary the third book giving $\binom{4}{1}$.

   $\therefore$ Total $= \binom{6}{3} - \binom{4}{1} = 16$.

   $\blacksquare$

Bucketing with Distinct Objects

DEFINITION 2.2.7 Bucketing Distinct Items: Suppose we would like to place n distinct items into r containers. The number of ways of doing so is:

$$r^n$$

This is a straight forward application of 2.1.1. Since there are n items, there are n steps. For each step there are r outcomes which represent the r container choices. Thus, the total number of outcomes or ways of placing n items into r containers is r^n.

Bucketing with In-Distinct Objects

DEFINITION 2.2.8 Divider Method (aka Bars and Stars): Suppose we would like to place n indistinct items into r containers. We imagine n star shapes ($\star$) to represent the indistinguishable items and $(r-1)$ bar shapes ($|$) to serve as the **container dividers**. Each unique configurations of this model represents one way to place n items into r containers. For example, one way to place $n = 6$ items into $r = 4$ containers is as follows:

$$\star \star \star | \star | \star | \star$$

In order to count every such configuration, we imagine $(n + r - 1)$ total "items", and find either n combinations of stars or $(r - 1)$ combinations of bars:

$$\binom{n + r - 1}{n} = \binom{n + r - 1}{r - 1} = \frac{(n + r - 1)!}{n!(r - 1)!}$$

or equivalently we can use 2.2.3 where we let n_1 be the number of indistinct stars and n_2 be the number of indistinct bars:

$$\frac{n!}{n_1! n_2!} = \frac{(n + r - 1)!}{n!(r - 1)!}$$

Example 2.2.9: Say we want to invest up to a maximum of \$10M into 4 companies. We can only invest in \$1M increments, and it is not necessary to invest everything.

Solution 2.2.9: We can introduce ourselves or our own bank account as a 5th "company" to account for situations where all of the money is not invested. This problem can now be modeled using $n = 10$ stars to represent the indistinguishable money "objects" and $r = 5$ to represent the company "containers". Thus, the total ways to invest in this case is:

$$\binom{10 + 5 - 1}{10} = \frac{(10 + 5 - 1)!}{10!(5 - 1)!} = 1001$$

■

Bucketing into Fixed Sized Containers

DEFINITION 2.2.10 Divider Method (aka Bars and Stars): If n objects are distinct, then the number of ways of putting them into r groups, such that group i has size n_i, and $\sum_{i=1}^{r} n_i = n$, is:

$$\frac{n!}{n_1! n_2! \cdots n_r!} = \binom{n}{n_1, n_2, \ldots, n_r}$$

where $\binom{n}{n_1, n_2, \ldots, n_r}$ is special notation called the multinomial coefficient.

Note that this is the exact same formula as 2.2.3. We can imagine this problem by replacing our n distinct items with n group "items" that are r unique. For each group i there would be n_i total group items all with number i. For example with $n_1 = 2$, $n_2 = 3$ and $n_3 = 1$ our object set would look like $[1, 1, 2, 2, 2, 3]$. We then use $\binom{n}{n_1, n_2, \ldots, n_r}$ to count the total unique configurations giving us all the ways in which the n distinct items can be placed in r fixed sized buckets.

Example 2.2.11:   Say we have 13 distinct servers that need to be assigned to 3 datacenters where datacenters A, B, and C have 6, 4, and 3 empty server racks, respectively. How many divisions of servers are possible?

Solution 2.2.11:   We model our object set as $[A, A, A, A, A, A, B, B, B, B, C, C, C]$ and find the number of permutations of these $n = 13$ items with $n_A = 6$, $n_B = 4$, and $n_C = 3$ indistinct objects:

$$\binom{n}{n_A, n_B, \ldots, n_C} = \binom{13}{6, 4, 3} = 60060$$

■

2.3   Definition of probability

DEFINITION 2.3.1 Sample Space: A Sample Space is the set of all possible outcomes of an experiment. For example:

- Experiment: Coin Flip, $S = \{Heads, Tails\}$

- Experiment: Flipping Two Coins, $S = \{(H, H), (H, T), (T, H), (T, T)\}$

- Experiment: Roll of 6-sided die, $S = \{1, 2, 3, 4, 5, 6\}$

- Experiment: # of emails received in a day, $S = \{x | x \in \mathbb{Z}, x \geq 0\}$

- Experiment: Youtube Hours in a Day, $S = \{x | x \in \mathbb{R}, 0 \leq x \leq 24\}$

DEFINITION 2.3.2 Sample Space: An event is some subset of the sample space to which we can ascribe meaning. Each of the following examples are read, "The event that ...":

- Event: a coin flip is heads, $E = \{Heads\}$

- Event: you flip at least 1 head on two coin flips, $E = \{(H, H), (H, T), (T, H)\}$

- Event: you roll 3 or less on a 6-sided die, $E = \{1, 2, 3\}$

- Event: you receive less than 20 emails in a day, $E = \{x | x \in \mathbb{Z}, 0 \leq x < 20\}$

- Event: you wasted ≥ 5 hours on Youtube, $E = \{x | x \in \mathbb{R}, 5 \leq x \leq 24\}$

Events are represented as capital letters, usually E or F, and are binary: they either happen or they don't.