Foundations of Machine Learning: Week 1: Terminology and Rules

Professor Christopher Yau
christopher.yau@manchester.ac.uk

October 21, 2020

2020-10-21

Foundations of Machine Learning:Week 1: Terminology and Rules

Foundations of Machine Learning: Week 1: Terminology and Rules

Professor Christopher Yau
christopher.yau@manchester.ac.uk
October 21, 2020

Sample space

A **sample space** (S) is the set of *all* possible outcomes of a random experiment.

Example

- Coin toss, $S = \{H, T\}$
- Die roll, $S = \{1, 2, 3, 4, 5, 6\}$
- Number of goals scored in a football match, $S = \mathbb{Z}^+$
- Output classes, $S = \{C_1, C_2, C_3, C_4\}$

Sample space

A **sample space** (S) is the set of *all* possible outcomes of a random experiment.

Example

- Coin toss, $S = \{H, T\}$
- Die roll, $S = \{1, 2, 3, 4, 5, 6\}$
- Number of goals scored in a football match, $S = \mathbb{Z}^+$
- Output classes, $S = \{C_1, C_2, C_3, C_4\}$

We will first start this lecture by talking about random experiments.
A *random experiment* is any experiment in which the outcome is uncertain (not known or determined in advance).
A *sample space* describes all the possible outcomes of a random experiment.
One of the surprisingly hardest challenges in real world machine learning problems can actually be to work out the sample space.
If this is not done properly, you may end up with scenarios which your algorithm cannot cater for!

Events

An **event** (E) is a subset of the sample space ($E \overbrace{\in}^{\text{in}} S$).

It is a subset of outcomes that might arise from a random experiment.

We denote the probability that an event E occurs using the notation $P(E)$.

Events

An **event** (E) is a subset of the sample space ($E \overbrace{\in}^{\text{in}} S$).
It is a subset of outcomes that might arise from a random experiment.
We denote the probability that an event E occurs using the notation $P(E)$.

In a random experiment, any subset of outcomes that could occur sis called an **event**.
Therefore each event describes a subset of the sample space.
We use the notation $P(E)$ to describe the probability that an event E occurs.

Events

Example

- Two coin tosses:

  $S = \{HH, HT, TH, HH\}$

  $E = \{\text{Head on first throw}\} = \{HT, HH\}$

- Die roll and coin toss:

  $S = \{1H, 2H, 3H, 4H, 5H, 6H, 1T, 2T, 3T, 4T, 5T, 6T\}$

  $E = \{\text{Head and die greater than 3}\} = \{4H, 5H, 6H\}$

Events

Example
- Two coin tosses:
$S = \{HH, HT, TH, HH\}$
$E = \{\text{Head on first throw}\} = \{HT, HH\}$
- Die roll and coin toss:
$S = \{1H, 2H, 3H, 4H, 5H, 6H, 1T, 2T, 3T, 4T, 5T, 6T\}$
$E = \{\text{Head and die greater than 3}\} = \{4H, 5H, 6H\}$

Here are two examples:
Suppose we have a coin tossing experiment involving two flips of the same coin.
The sample space describes every combination of heads and tails for that experiment.
An event might be the subset of outcomes where a head appears on the first flip.
Similarly, we could roll a six-sided die and flip a coin.
The outcomes are every combination of possible die roll and coin flip that could occur.
An event might be the subset of outcomes where we get a head and a die value greater than 3.

Events

The **empty set** or null event ($\emptyset$).

The sample space (S) is of course itself an event.

Events

The **empty set** or null event ($\emptyset$).
The sample space (S) is of course itself an event.

There are a couple of special events.
The empty set or null event is an event in which nothing has occurred.
While the sample space is itself an event (i.e.the event which includes all possible outcomes).
Use of these special events is sometimes necessary for deriving mathematical relationships.

Axioms

The probabilities of any event between 0 and 1 (inclusive)

$$0 \leq P(E) \leq 1$$

The outcome must be in the sample space S.

Therefore the event that consists of all possible outcomes (i.e. the sample space) has probability $P(S) = 1$

Axioms

The probabilities of any event between 0 and 1 (inclusive)

$0 \leq P(E) \leq 1$

The outcome must be in the sample space S.

Therefore the event that consists of all possible outcomes (i.e. the sample space) has probability $P(S) = 1$

Axioms describe the basic properties and rules by which we manipulate probabilities.
We will now consider some basic axioms.
First, the probability of an event must be between 0 and 1 by definition where 0 denotes no probability that the event could occur and 1 that the event is certain to occur.
Of course, by definition all events must be contained in the sample space.
Therefore the event of all possible outcomes, the sample space itself, must add up to 1.

Addition Law

Probability statements need to be made carefully before we can start combining them and making summary statements:

- Event E occurs **AND** event F occurs (**Joint**). $P(E \cap F)$
- Event E occurs **OR** event F occurs (**Addition Law**). $P(E \cup F)$
- Event E occurs **AND** event F **OR** event G occurs. $P(E \cap (F \cup G))$

Addition Law

Probability statements need to be made carefully before we can start combining them and making summary statements:

- Event E occurs **AND** event F occurs (**Joint**). $P(E \cap F)$
- Event E occurs **OR** event F occurs (**Addition Law**). $P(E \cup F)$
- Event E occurs **AND** event F **OR** event G occurs. $P(E \cap (F \cup G))$

The next axiom known as the addition law concerns how we add up probabilities.
When we have two or more events, we may wish to "add up" probabilities associated with these events to obtain a total.
We need to think carefully though about what we mean by "adding up" when dealing with probabilities.
A useful way of thinking is to think in terms of AND and OR operations.
For example:
What is the probability that event E occurs AND event F occurs? This is known as the joint probability.
What is the probability that event E occurs OR event F occurs? This is where we can apply the addition law that I will discuss shortly.
What is the probability that event E occurs AND event F OR event occurs? Involves combining joint and addition rules.
The addition law is concerned with OR operations.

Addition Law

If two events E and F are disjoint (they do not overlap)

$$E \overbrace{\cap}^{\text{AND}} F = \emptyset$$

e.g. rolling die, $E = \{2, 4, 6\}$, $F = \{1, 3, 5\}$ are disjoint

e.g. rolling die, $E = \{1, 2\}$, $F = \{2, 4, 6\}$ is not disjoint

Then the addition law says:

$$P(E \underbrace{\cup}_{\text{OR}} F) = P(E) + P(F)$$

Addition Law

If two events E and F are disjoint (they do not overlap)

$$E \overbrace{\cap}^{\text{AND}} F = \emptyset$$

e.g. rolling die, $E = \{2, 4, 6\}$, $F = \{1, 3, 5\}$ are disjoint

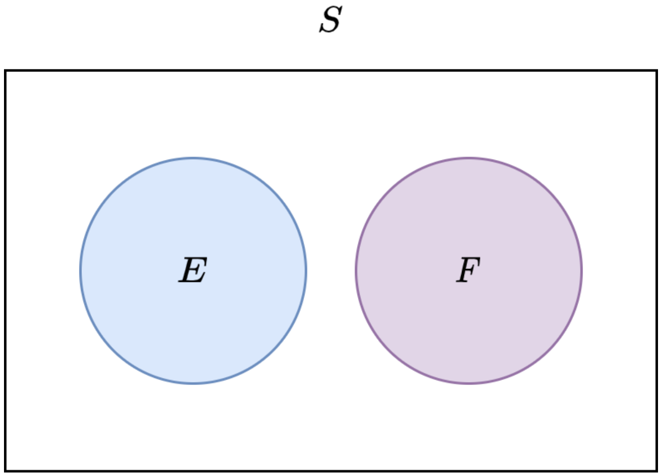
e.g. rolling die, $E = \{1, 2\}$, $F = \{2, 4, 6\}$ is not disjoint

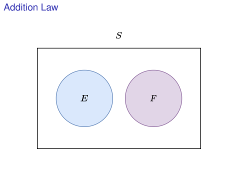
Then the addition law says:

$$P(E \underbrace{\cup}_{\text{OR}} F) = P(E) + P(F)$$

If we want the total probability of two events we need to be concerned with whether the events overlap in terms of outcomes.
If two events are *disjoint* this means the two events do not overlap and therefore it is not possible for both things to happen at the same time.
For example, when rolling a die, the events for getting an even number and getting an odd number are disjoint.
However, the events for getting a number less than three and an even number are not disjoint.
The addition law says when two events are joint, this means the probability of one event *or* another occurring is simply the sum of the two individual events.

Addition Law





We can illustrate this using a Venn diagram.
Here the box represents the sample space containing all possible outcomes.
The two circles are our events.
In this case the two circles do not overlap, the events are disjoint so we can add the probabilities to get the total that either event occurs.

If the events did overlap, we need to take that overlap into special consideration to avoid *double counting*.

Addition Law

The **addition law** of probability for non-disjoint events:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

The probability that E or F occurs is the probability of E and F minus the probability that *both* occur (to avoid double counting).

Addition Law

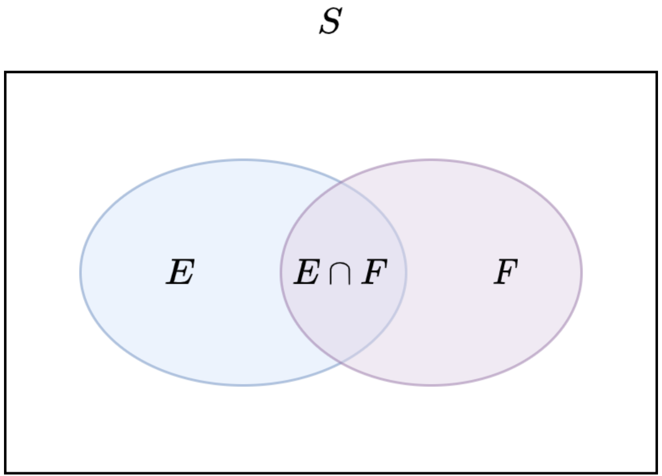
The **addition law** of probability for non-disjoint events:

$P(E \cup F) = P(E) + P(F) - P(E \cap F)$

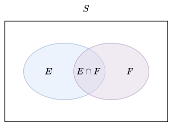
The probability that E or F occurs is the probability of E and F minus the probability that *both* occur (to avoid double counting).

In general, if two events are not disjoint, we need to “subtract” the overlapping region to avoid double counting outcomes contained within both events.
So the addition law of probability has the following form that the probability of event E or F occurring is equal to the sum of the individual event probabilities minus the probability that they both occur.

Addition Law



Addition Law



This Venn diagram illustrates the overlap.
When we add the probabilities of the two events E and F we double count the region in the middle so we need to take one copy of it away.

Complement Law

The **complement law** states that:

$$P(E) = 1 - P(\overbrace{\overline{E}}^{\text{Not E}})$$

The probability of the event E is one minus the probability that event E does not occur.

Note - $P(S) = P(E) + P(\overline{E}) = 1$

Complement Law

The **complement law** states that:

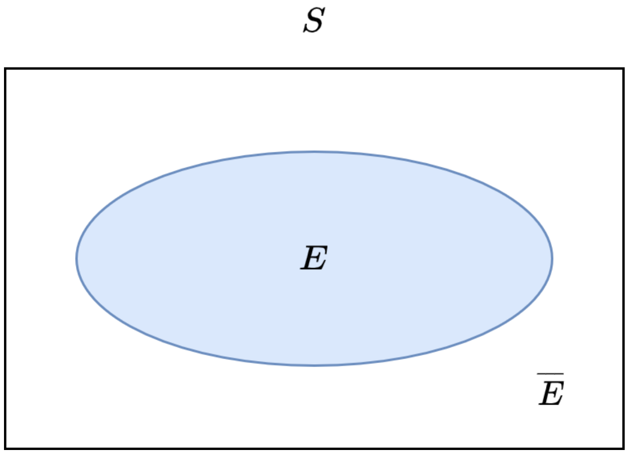
$$P(E) = 1 - P(\overbrace{\overline{E}}^{\text{Not E}})$$

The probability of the event E is one minus the probability that event E does not occur.

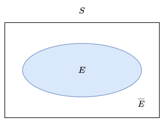
Note - $P(S) = P(E) + P(\overline{E}) = 1$

The next axiom known as the complement law concerns how we deal with opposites.
The complement law tells us that the probability of an event happening is 1 minus the probability that it doesn't occur or vice-versa.
This is because if all outcomes have been considered in our sample space, the probabilities of all outcomes must add up to 1.
Therefore the probability of any event and its complement or negative must add up to 1.

Complement Law



Complement Law



This is illustrated in this Venn diagram where the event E and its complement $\tilde{E}$ which is everything not in E must by definition form the entire sample space.

Complement Law

The probability that there is *more than* two car accidents in a day is one minus the probability that there are *at least* two car accidents in a day:

$$
\begin{aligned}
P(\text{Accident} > 2) = 1 &- P(\text{Accident} = 2) \\
&- P(\text{Accident} = 1) \\
&- P(\text{Accident} = 0)
\end{aligned}
$$

In numerical problems, it might be easier to solve the negated problem and apply the complement law to obtain the desired answer.

Complement Law

The probability that there is *more than* two car accidents in a day is one minus the probability that there are *at least* two car accidents in a day:

$$
\begin{aligned}
P(\text{Accident} > 2) = 1 &- P(\text{Accident} = 2) \\
&- P(\text{Accident} = 1) \\
&- P(\text{Accident} = 0)
\end{aligned}
$$

In numerical problems, it might be easier to solve the negated problem and apply the complement law to obtain the desired answer.

The complement law can be very useful for certain computations where one wishes to compute the probability of an event but it is actually easy to compute the probability of that event not occurring and then doing one minus and the complement law to get the desired answer.
For example, what is the probability that there are more than two car accidents in a day?

We can compute this as one minus the probability of 0, 1 and 2 accidents rather than trying to enumerate an infinite number of possible accidents.

Numerical Example

A card is selected at random from a normal deck of cards.

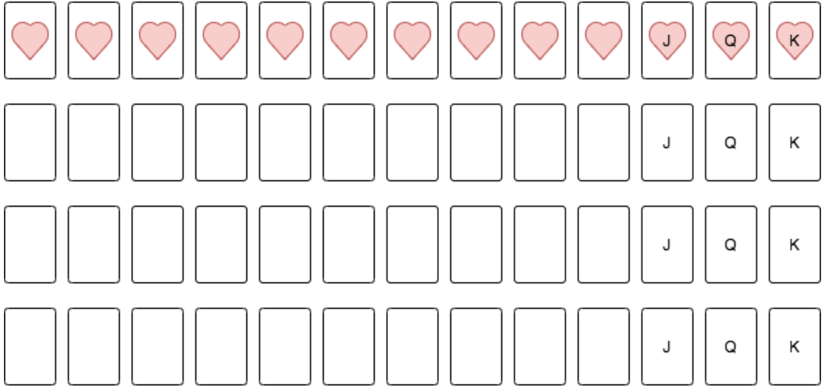
What is the probability of drawing a heart or a suit card?

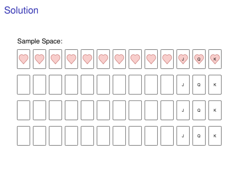
A card is selected at random from a normal deck of cards.
What is the probability of drawing a heart or a suit card?

Let us look at a simple example ...
A card is selected at random from a normal deck of cards.
What is the probability of drawing a heart or a suit card?

Solution

Sample Space:

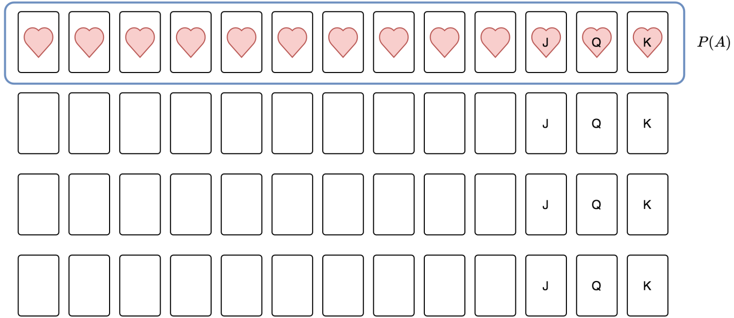


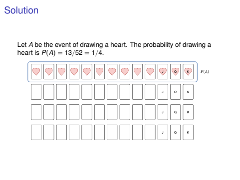


We first need to identify the sample space which consists of 52 cards, 13 hearts, and 12 suit cards (4 jacks, 4 queens and 4 kings).

Solution

Let A be the event of drawing a heart. The probability of drawing a heart is $P(A) = 13/52 = 1/4$.

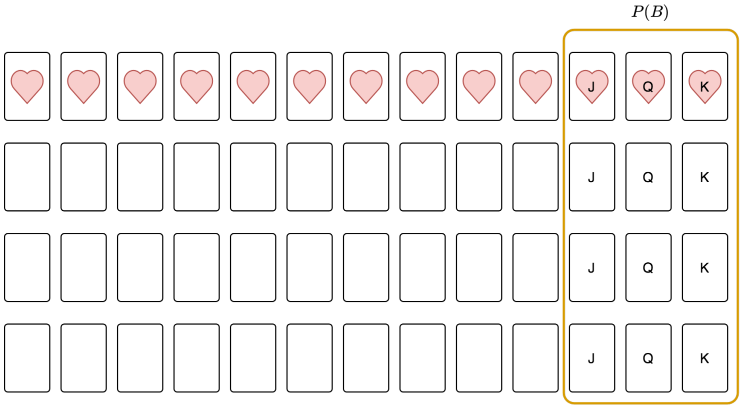


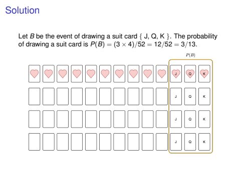


Now we need to denote the event of drawing a heart, lets call that event A. There are 13 hearts out of 52 cards, so the probability is 13/52 or 1/4.

Solution

Let B be the event of drawing a suit card { J, Q, K }. The probability of drawing a suit card is $P(B) = (3 \times 4)/52 = 12/52 = 3/13$.



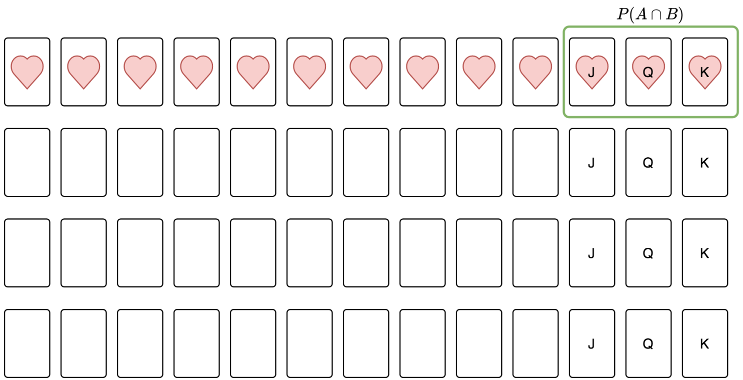


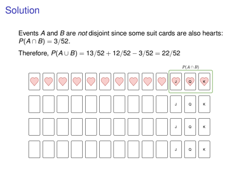
Now we need to denote the event of drawing a suit card, lets call that event B. There are 12 suit cards out of 52 cards, so the probability is 12/52 or 3/13.

Solution

Events A and B are *not* disjoint since some suit cards are also hearts: $P(A \cap B) = 3/52$.

Therefore, $P(A \cup B) = 13/52 + 12/52 - 3/52 = 22/52$





Notice that the two events are *not* disjoint since some suit cards are also hearts. There are 3 out of 52 such cards.
So using our addition law, we can compute the probability of a heart or suit card being drawn as the probability of a heart (13/52) plus the probabiility of a suit card (12/52) minus the probability of a card being both a heart and suit card (3/52) to give 22/52 as the final answer.

Recap: Probability

Sample space (S) describes *all* possible outcomes of an experiment.

Event ($E \in S$) describes a *subset* of possible outcomes.

Recap: Probability

Sample space (S) describes *all* possible outcomes of an experiment.
Event ($E \in S$) describes a *subset* of possible outcomes.

So to recap, the sample space describes all possible outcomes of a random experiments, while events describes subsets of possible outcomes.

Recap: Probability

The *joint* probability that both events E **and** F occurs is $P(E \cap F)$.

The probability that event E **or** F occurs is:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Two events E and F are **disjoint** if:

$$P(E \cap F) = 0.$$

Recap: Probability

The *joint* probability that both events E **and** F occurs is $P(E \cap F)$.
The probability that event E **or** F occurs is:
$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$
Two events E and F are **disjoint** if:
$$P(E \cap F) = 0.$$

We need to be able to differentiate between joint probability statements, event E AND F happening from OR events in which E or F can occur.
For OR events, the addition law tells us how to find these probabilities and the form of the addition law simplifies if the events are disjoint.

END LECTURE

END LECTURE