Contents

1 Naive Bayes

1.1 1. Overarching Goals

- We want to assign each $\bar{x} \in X$ to its correct class C_k, $k \in 1, ..., K$
- We ask: **Given** $\bar{x} \in X$, what is its most **probable class** C_k?

1.2 Understanding Conditional Probability

- Sample space (Ω) contains events whose probability of occurring is proportional to its area w.r.t the sample space
- In a fair coin flip, $\mathbb{P}(heads) = \mathbb{P}(H) = \mathbb{P}(tails) = \mathbb{P}(T) = 0.5$
- Say we have the events H, T, and T, where H is the event that Hillary Clinton is elected president, T is the probability that Trump is, and N is the probability that a nuclear winter occurs during the next presidency.
- see figure one
- If we want to talk about the probability of a nuclear winter happening ($\mathbb{P}(N)$) given that Donald Trump is elected president ($\mathbb{P}(T)$), we express it as $\mathbb{P}(N|T)$.
  - $\mathbb{P}(N|T)$ is read as "probability of a nuclear winter given that Trump is elected"
- $\mathbb{P}(N|T) = \frac{\mathbb{P}(N \cap T)}{\mathbb{P}(T)}$.
  - Note that the probability of event T occurring is in the denominator, think of this as the new sample space that has been reduced.
  - We then find the probability of $\mathbb{P}(N \cap T)$ over this new, reduced sample space. This is equivalent to $\mathbb{P}(N|T) = \frac{\mathbb{P}(N \cap T)}{\mathbb{P}(T)}$.
  - see figure 2

1.3 Bayes' Theorem

- We can represent the entire sample space Ω in terms of conditional probabilities.
- see figure 3
- As a reminder: we want to assign each $\bar{x} \in X$ to its correct class C_k, $k \in 1, ..., K$
- We ask: **Given** $\bar{x} \in X$, what is its most **probable class** C_k?
- $P(C_k|\bar{x}) = \frac{P(\bar{x}|C_k) \cdot P(C_k)}{P(\bar{x})}$
- $P(C_k|\bar{x}) = \frac{P(C_k \cap \bar{x})}{P(\bar{x})} = \frac{P(C_k, \bar{x})}{P(\bar{x})}$
  - $= \frac{P(x_1, x_2, ..., x_n, C_k)}{P(\bar{x})}$

- $= \frac{P(x_1|x_2,\ldots,x_n,C_k)\cdot P(x_2,\ldots,x_n,C_n)}{P(\bar{x})}$
- $= \frac{P(x_1|x_2,\ldots,C_k)\cdot P(x_2|x_3,\ldots,C_k)\ldots P(x_n|C_k)P(C_k)}{P(\bar{x})}$

- See figure 4.

1.4 Terminology / Glossary

- Sample Space (Ω) - set of all possible outcomes of an experiment
- A, B - events that are subsets of the sample space
- $\mathbb{P}(A)$ - Probability of event A occurring.
- $\mathbb{P}(-A)$ (or $\mathbb{P}(A^c)$) - Probability of $A's$ complement ("not A") occurring.
- $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\cdot\mathbb{P}(B|A)}{\mathbb{P}(B)}$
- $A \perp B$ - A is independent of B if and only if the (non-)occurrence of B has no effect on the (non-)occurrence of A, and vice versa.
  - Corollary: If $A \perp B$, $\mathbb{P}(A|B) = \mathbb{P}(A)$
  - Technical definition: $A \perp B$ if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

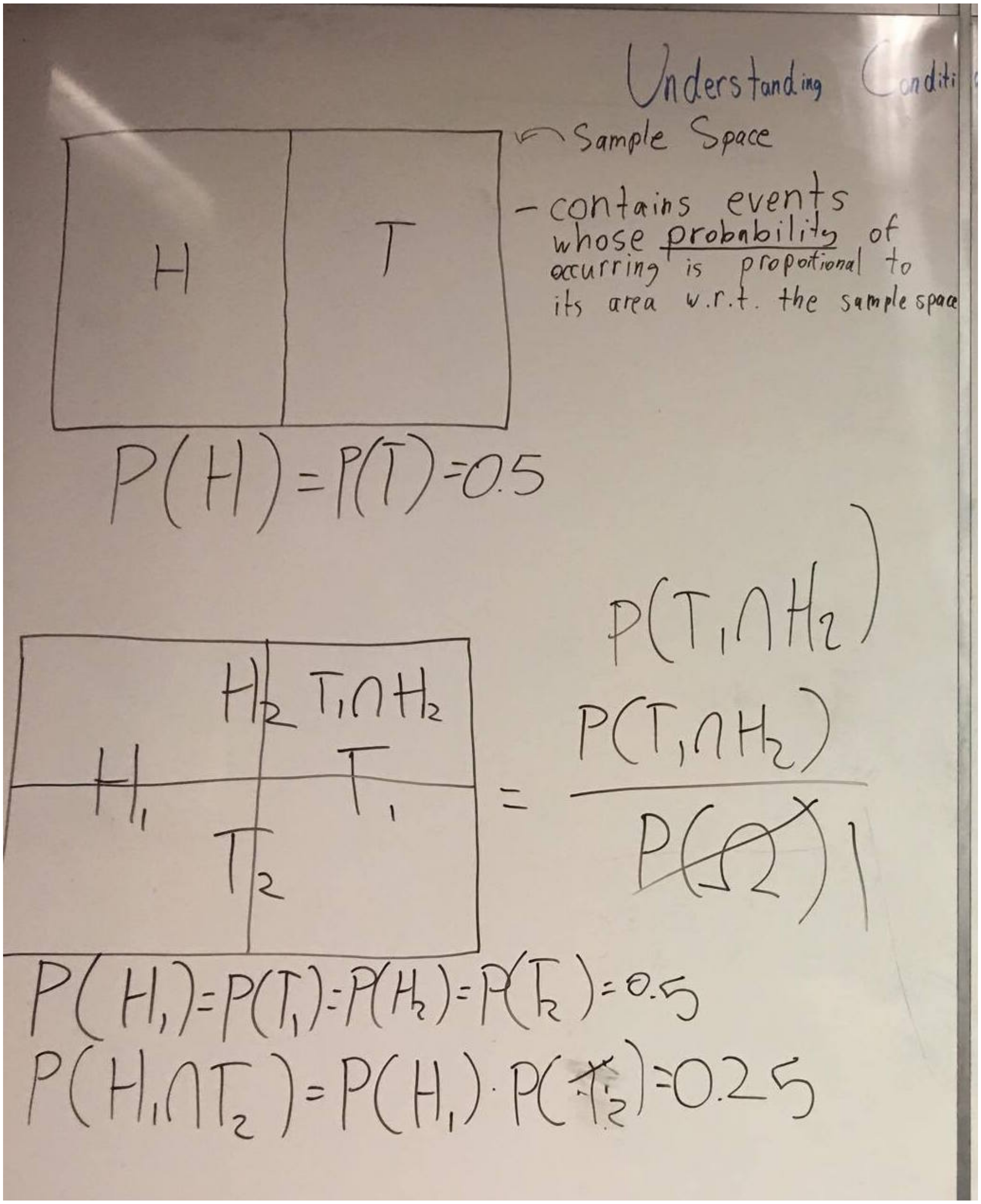

Figure 1: A visual representation of the sample space in the case of one and two coin flips.

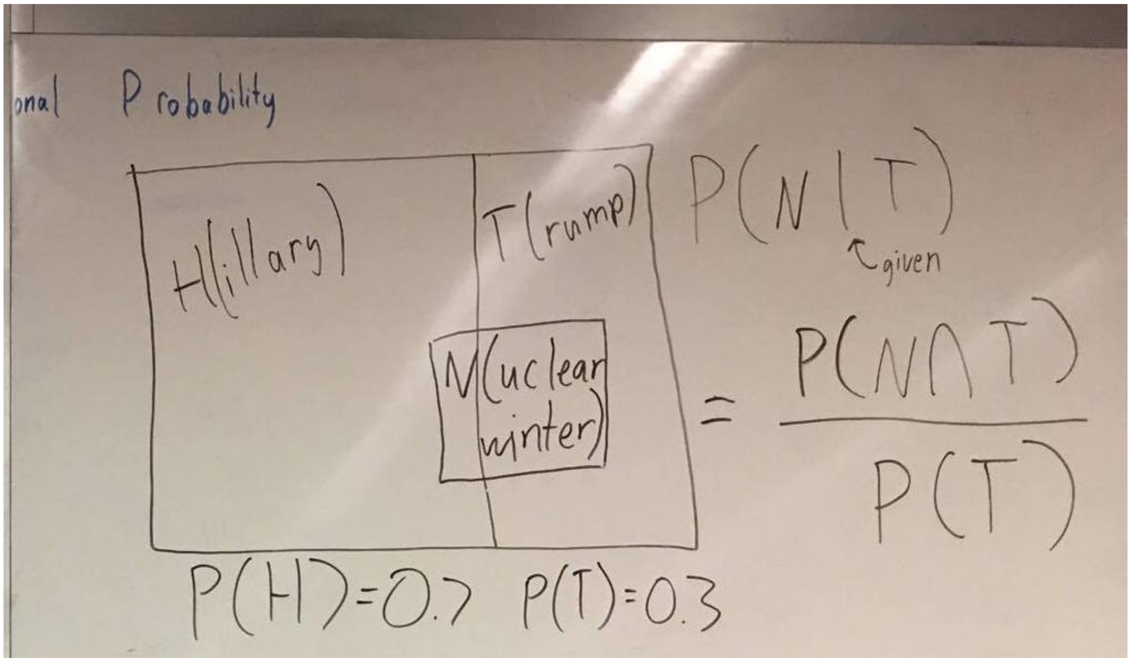

Figure 2: A visual representation of the sample space in conditional probability.

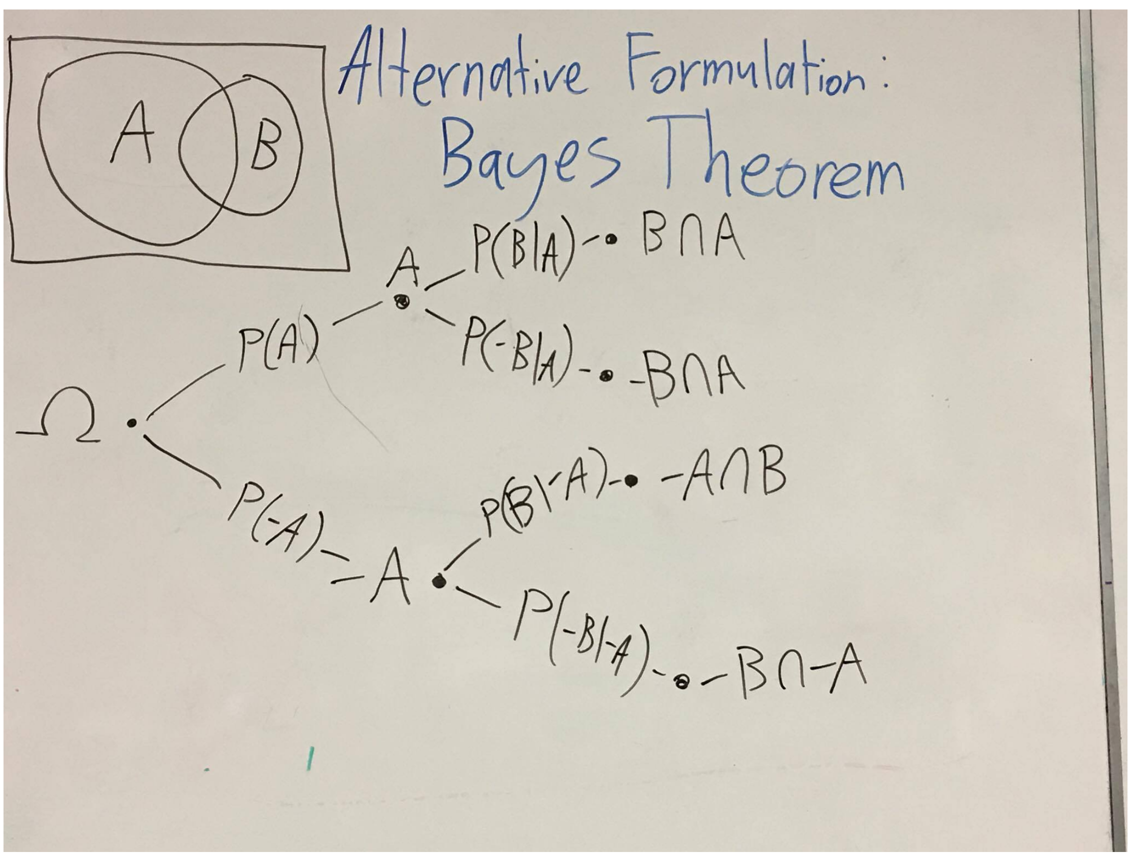

Figure 3: A visual representation of the sample space as conditional probabilities

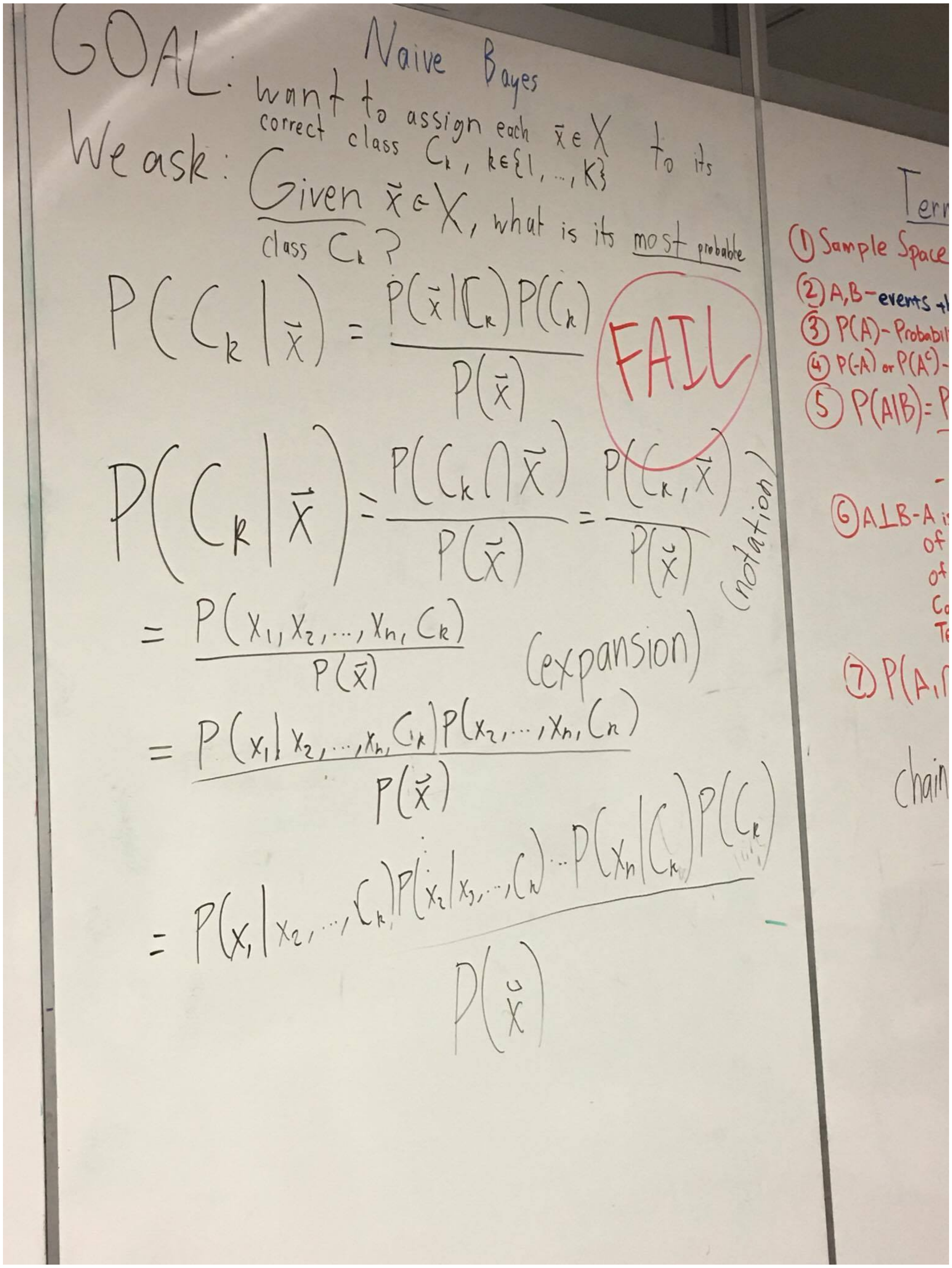

Figure 4: The expansion of bayes' theorem