CS70 - Lecture 17 Notes

Name: Felix Su   SID: 25794773

Spring 2016   GSI: Gerald Zhang

Review

- Example: $B \subset A \Rightarrow A$ and B are positively correlated
  - $\Pr[A|B] = 1 > \Pr[A]$ and $\Pr[A \cap B] = \Pr[B] > \Pr[A]\Pr[B]$
- Example: $A \subset B = \emptyset \Rightarrow A$ and B are negatively correlated
  - $\Pr[A|B] = 0 < \Pr[A]$ and $\Pr[A \cap B] = 0 < \Pr[A]\Pr[B]$
- For uniformly distributed probability space Ω, $\Pr[A] = \frac{|A|}{|\Omega|}$

Probability of A given B:

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} \tag{1}$$

Probability of A and B (intersection):

$$\Pr[A \cap B] = \Pr[B]\Pr[A|B] = \Pr[A]\Pr[B|A] \tag{2}$$

A and B are positively correlated if:

$$\Pr[A|B] > \Pr[A] \quad , \quad \Pr[A \cap B] > \Pr[A]\Pr[B] \tag{3}$$

A and B are negatively correlated if:

$$\Pr[A|B] < \Pr[A] \quad , \quad \Pr[A \cap B] < \Pr[A]\Pr[B] \tag{4}$$

A and B are independent iff:

$$\Pr[A|B] = \Pr[A] \quad , \quad \Pr[A \cap B] = \Pr[A]\Pr[B] \tag{5}$$

Find prior probability given some observation B (A given B)

1. Total probability of B given prior probabilities
   - **Law of Total probability**
   - $\Pr[B] = \Pr[A_1]\Pr[B|A_1] + \cdots + \Pr[A_n]\Pr[B|A_n]$
2. Find $\Pr[A|B]$
   - **Bayes Rule**
   - $\Pr[A|B] = \frac{\Pr[A]\Pr[B|A]}{\Pr[B]}$

Terms

- **Most likely A Posteriori (MAP) of** B: The A_m that gives the highest $\Pr[A_m]\Pr[B|A_m]$
- **Maximum Likelihood Estimate (MLE) of** B: The A_m that gives the highest $\Pr[B|A_m]$

Mutual Independence

- A subset of events $A_1, ..., A_k$ where $A_k, k \in J$ are **mutually independent** if the probability that they all occur is equal to the product of their individual probabilities

Mutual Independence Definition

$$\Pr[\cap_{k\in K} A_k] = \prod_{k\in K} \Pr[A_k], \text{ for all finite } K \subseteq J \tag{6}$$

Theorem

- If the events $\{A_j, j \in J\}$ are mutually independent, and if K_n are pairwise disjoint finite subsets of J, then all the events $\cap_{k\in K_n} A_k$ are independent (same is true if we replace some of the A_k by $\bar{A}_k$

Collision Calculation
Let m = no. of elements, n= no. of bins, C = collision

$$\Pr[\bar{C}] \approx e^{(-\frac{m^2}{2n})} \tag{7}$$

When $m = 1.2\sqrt{n}$

$$\Pr[C] \approx \frac{1}{2} \tag{8}$$

Collision Derivation
If A_i = no collision when the ith ball is placed in a bin

$$\Pr[A_i | A_{i-1} \cap \cdots \cap A_1] = 1 - \frac{i-1}{n} \tag{9}$$

No collisions = $A_1 \cap \cdots \cap A_m$
Product Rule:

$$\Pr[A_1 \cap \cdots \cap A_m] = \Pr[A_1]\Pr[A_2|A_1] \cdots \Pr[A_m|A_1 \cap \cdots \cap A_{m-1}] \tag{10}$$

Apply to $\Pr[\bar{C}]$:

$$\Pr[\bar{C}] = (1 - \frac{1}{n}) \cdots (1 - \frac{m-1}{n}) \tag{11}$$

Natural log of both sides:

$$\ln(\Pr[\bar{C}]) = \sum_{k=1}^{m-1} \ln(1 - \frac{k}{n}) \approx \sum_{k=1}^{m-1} \ln(-\frac{k}{n})^{*} = (-\frac{1}{n})(\frac{m(m-1)}{2}) \approx -\frac{m^2}{2n} \tag{12}$$

* Use property that $\ln(1-\varepsilon) \approx -\varepsilon$ for $|\varepsilon| << 1$
Gauss Summation: $1 + 2 + \cdots + m - 1 = \frac{m(m-1)}{2}$

Example: Checksums

- m = no. of files, b = no. of bits in the checksum, C = files share a checksum
- Find b s.t. $\Pr[C] \leq 10^{-3}$
  - $\Pr[C] \approx 1 - e^{(-\frac{m^2}{2(2^b)})}$
  - $b = \frac{\ln(-\frac{m^2}{2\ln(1-10^{-3})})}{\ln(2)} = 2.9\ln(m) + 9$
- $\therefore b \geq 2.9\ln(m) + 9$

Probability of Getting n_i out of n with m picks

- Define event of failure A_m (not success)
- Determine probability of failing on each iteration of m
  - $\Pr[A_i|A_{i-1} \cap \cdots \cap A_1] = 1 - \Pr[\bar{A}_i]$ for $i = \{1, ..., m\})$
  - If not intuitive, try brute force and find a pattern for each $\Pr[A_i]$
- Use Product Rule to get $\Pr[A_m]$
  - $\Pr[A_1 \cap \cdots \cap A_m] = \Pr[A_1]\Pr[A_2|A_1] \cdots \Pr[A_m|A_1 \cap \cdots \cap A_{m-1}]$
  - If events are **independent** $\Pr[A_1 \cap \cdots \cap A_m] = \Pr[A_1]\Pr[A_2] \cdots \Pr[A_m|A_{m-1}]$
- Take natural log of both sides and simplify using the property that $\ln(1 - \varepsilon) \approx -\varepsilon$ for $|\varepsilon| << 1$
- Raise e to the power of both sides (e^n) to derive approximate solution for $\Pr[A_m]$
  - $\Pr[A_m] \approx e^{expression}$

Probability of Complete Collection

- Define event of failure of one iteration E_k
  - E_k for $k = \{1, ..., n\}$
  - Derive $\Pr[E_k]$ using method above: **Probability of Getting n_i out of n with m picks**
- find probability of failing any iteration (or/union)
  - $p := \Pr[E_1 \cup E_2 \cup \cdots \cup E_n]$
- Estimate p using Union Bound
  - $p := \Pr[E_1 \cup E_2 \cup \cdots \cup E_n] \leq \Pr[E_1] + \Pr[E_2] + \cdots + \Pr[E_n]$
- Plug in $\Pr[E_k]$ expression derived above to find $\Pr[$ failure of at least one iteration $] \leq expression$
- Use expression to derive minimum value of m to get a certeain $\Pr[miss]$ s.t. $\Pr[miss] \leq p$