Independence

Independence is a big deal for machine learning. The joint probability of many events requires exponential amounts of data. By making independence and conditional independence claims, computers can calculate accurate probabilities very fast, based on a lot less data.

Independence

Two events E and F are called independent if: $P(EF) = P(E) P(F)$. Or, equivalently: $P(E \mid F) = P(E)$. Otherwise, they are called dependent events

Three events E, F, and G independent if:
$P(EFG) = P(E) P(F) P(G)$, and
$P(EF) = P(E) P(F)$, and
$P(EG) = P(E) P(G)$, and
$P(FG) = P(F) P(G)$

If events E and F and independent, E^C and F^C are also independent. Independence unfortunately is not transitive.

Generally n events E_1, E_2, ..., E_n are independent if for every subset with r elements (where $r \leq n$) it holds that: $P(E_a, E_b, ... E_r) = P(E_a)P(E_b)P(E_r)$. For example: the outcomes of n separate flips of a coin are all independent of one another. Each flip in this case is called a "trial" of the experiment.

Hash Map Example

Let's consider our friend the hash map. We are going to hash m strings (unequally) into a hash table with n buckets. Each string hashed is an independent trial, with probability p_i of getting hashed to bucket i. Calculate the probability of these three events:

A) E = *at least one string hashed to first bucket*
B) E = <u>At least 1 of</u> buckets 1 to k has ≥ 1 string hashed to it
C) E = <u>Each of</u> buckets 1 to k has ≥ 1 string hashed to it

Part A

Let F_i be the event that string i is not hashed into the first bucket. Note that all F_i are independent of one another. By mutual exclusion $P(F_i) = (p_2 + p_3 + ... + p_n)$

$P(E)$	$= 1 - P(E^C)$	Since $P(A) + (A^C) = 1$
	$= 1 - P(F_1 F_2 ... F_m)$	By the semantics of F_i
	$= 1 - P(F_1)P(F_2)...P(F_m)$	Since the events are independent
	$= 1 - (p_2 + p_3 + ... + p_n)^m$	We calculate $P(F_i)$ by mutual exclusion.

Part B

Let F_i be the event that at least one string is hashed into bucket i. Note that F_is are neither independent nor mutually exclusive.

$$
\begin{aligned}
P(E) \quad &= P(F_1 \cup F_2 \cup ... \cup F_k) \\
&= 1 - P([F_1 \cup F_2 \cup ... \cup F_k]^C) \qquad \text{Since } P(A) + (A^C) = 1 \\
&= 1 - P(F_1{}^C F_2{}^C ... F_k{}^C) \qquad \text{By DeMorgan's Law} \\
&= 1 - (1 - p_1 - p_2 - ... - p_k)^m \qquad \text{By thinking about the semantics}
\end{aligned}
$$

The last step is calculated by realizing that $P(F_1{}^C F_1{}^C ... F_k{}^C)$ is only satisfied by m independent hashes into buckets other than 1 through k.

Part C

Let F_i be same as in Part B.

$$
\begin{aligned}
P(E) \quad &= P(F_1 F_2 ... F_k) \\
&= 1 - P([F_1 F_2 ... F_k]^C) \qquad \text{Since } P(A) + (A^C) = 1 \\
&= 1 - P(F_1{}^C \cup F_2{}^C \cup ... \cup F_k{}^C) \qquad \text{By DeMorgan's Other Law} \\
&= 1 - P\left(\bigcup_{i=1}^{k} F_i^c\right) = 1 - \sum_{r=1}^{k} (-1)^{(r+1)} \sum_{i_1 < ... < i_r} P(F_{i_1}{}^c F_{i_2}{}^c ... F_{i_r}{}^c) \qquad \text{By Inclusion/Exclusion}
\end{aligned}
$$

Where $P(F_1{}^C F_1{}^C ... F_k{}^C) = (1 - p_1 - p_2 - ... - p_k)^m$ just like in the last problem.

Conditional Independence

Two events E and F are called conditionally independent given G, if
P(E F | G) = P(E | G) P(F | G)
Or, equivalently: *P(E | F G) = P(E | G)*

Conditional Breaking Independence

Let's say a person has a fever if they either have malaria or have an infection. We are going to assume that getting malaria and having an infection are independent: knowing if a person has malaria does not tell us if they have an infection. Now, a patient walks into a hospital with a fever. Your belief that the patient has malaria is high and your belief that the patient has an infection is high. Both explain why the patient has a fever.

Now, given our knowledge that the patient has a fever, gaining the knowledge that the patient has malaria **will** change your belief the patient has an infection. The malaria explains why the patient has a fever, and so the alternate explanation becomes less likely. The two events (which were previously independent) are dependent when conditioned on the patient having a fever..

Conditioning on an event E leads to dependence between previously independent events A and B when A and B are independent causes of E.

Disclaimer: This handout was made fresh just for you. Did you notice any mistakes? Let Chris know and he will fix them.