BAYES' THEOREM FOR RANDOM VARIABLES: A COMPUTATIONAL VIEW

Computationally, Bayes' theorem can be thought of as a two-step procedure. Once we have observed $Y = y$:

1. For each value x that random variable X can take on, initially we believed that $X = x$ with a score of $p_X(x)$, which could be thought of as how plausible we thought ahead of time that $X = x$. However now that we have observed $Y = y$, we weight the score $p_X(x)$ by a factor $p_{Y|X}(y \mid x)$, so

$$\text{new belief for how plausible } X = x \text{ is:} \quad \alpha(x \mid y) \triangleq p_X(x) p_{Y|X}(y \mid x), \tag{2.4}$$

   where we have defined a new table $\alpha(\cdot \mid y)$ which is *not* a probability table, since when we put in the weights, the new beliefs are no longer guaranteed to sum to 1 (i.e., $\sum_x \alpha(x \mid y)$ might not equal 1)! $\alpha(\cdot \mid y)$ is an *unnormalized* posterior distribution!

   Also, if $p_X(x)$ is already 0, then as we already mentioned a few times, $p_{Y|X}(y \mid x)$ is undefined, but this case isn't a problem: no weighting is needed since an impossible outcome stays impossible.

   To make things concrete, here is an example from the medical diagnosis problem where we observe $Y = \text{positive}$:

Homework due Oct 05, 2016 at 21:00 UTC

Notation Summary (Up Through Week 3)

Mini-project 1: Movie Recommendations (Week 3)
Mini-projects due Oct 12, 2016 at 21:00 UTC

p_X

healthy	infected
0.999	0.001

$p_{Y|X}$

Y \ X	healthy	infected
positive	0.01	0.99
negative	0.99	0.01

entry-wise multiply to get *unnormalized* posterior

healthy	infected
0.00999	0.00099

2. We fix the fact that the unnormalized posterior table $\alpha(\cdot \mid y)$ isn't guaranteed to sum to 1 by renormalizing:

$$p_{X|Y}(x \mid y) = \frac{\alpha(x \mid y)}{\sum_{x'} \alpha(x' \mid y)} = \frac{p_X(x)p_{Y|X}(y \mid x)}{\sum_{x'} p_X(x')p_{Y|X}(y \mid x')}.$$

An important note: Some times we won't actually care about doing this second renormalization step because we will only be interested in what value that X takes on is more plausible relative to others; while we could always do the renormalization, if we just want to see which value of x yields the highest entry in the unnormalized table $\alpha(\cdot \mid y)$, we could find this value of x without renormalizing!