EECS 445: Machine Learning **Fall 2016**

Discussion 9: Explaining Away

Written by: Benjamin R. Bray and Chansoo Lee

9.1 Probabilistic influence

Recall the HW5 Problem 3.

Lemma 9.1. $P(t_1|d_1) < P(t_1|d_0)$

Intuitively, this statement is kind of obvious. The person is on good diet (d_1) should be less likely to test for high cholesterol (t_1)

Proof: From the factorization theorem, T is independent of all variables other than C given D. So,

$$P(T|D, u) = P(T|D) \tag{9.1}$$

for all values of u.

Now from the definition, we have that

$$P(c_1|d_1) < P(c_1|d_0) \tag{9.2}$$

meaning people with good diet is less likely to have a high cholesterol than those with bad diet. We were able to eliminate the conditioning on u because of (9.1). Similarly,

$$P(t_1|c_1) > P(t_1|c_0).$$

Because T and D are independent given C, we can write

$$P(T|D) = \sum_c P(T, C = c|D) = \sum_c P(T|C = c)P(C = c|D).$$

Now,

$$P(t_1|d_1) = P(t_1|c_1)P(c_1|d_1) + P(t_1|c_0)P(c_0|d_1)$$

and

$$P(t_1|d_0) = P(t_1|c_1)P(c_1|d_0) + P(t_1|c_0)P(c_0|d_0)$$

We compute the difference:

$$\begin{aligned} P(t_1|d_0) - P(t_1|d_1) &= P(t_1|c_1)(P(c_1|d_0) - P(c_1|d_1)) + P(t_1|c_0)(P(c_0|d_0) - P(c_0|d_1)) \\ &= P(t_1|c_1)(P(c_1|d_0) - P(c_1|d_1)) + P(t_1|c_0)\big(1 - P(c_1|d_0) - (1 - P(c_1|d_1))\big) \\ &= \big(P(t_1|c_1) - P(t_1|c_0)\big)\big(P(c_1|d_0) - P(c_1|d_1)\big) \end{aligned}$$

which is positive because both terms in the parentheses are positive. ■

Corollary 9.2.

$$P(t_1|d_1) < P(t_1)$$

Proof: Note that

$$P(t_1) = P(t_1|d_1)p(d_1) + p(t_1|d_0)p(d_0) > P(t_1|d_1)p(d_1) + p(t_1|d_1)p(d_0) = P(t_1|d_1)$$ ■

9.2 Explaining Away

The explaining away is the following phenomenon:

$$P(h_1|b_0, e_1) < P(h_1|b_1, e_1). \tag{9.3}$$

It follows that

$$P(h_1|b_0, e_1) < P(h_1|e_1) < P(h_1|b_1, e_1).$$

Let's explore the sufficient and necessary condition for this to happen.

By the Bayes rule,

$$P(h_1|e_1, B) = P(e_1|h_1, B)\frac{P(h_1|B)}{P(e_1|B)} = P(e_1|h_1, B)\frac{P(h_1)}{P(e_1|B)}$$

since H and B are independent.

Note that

$$P(e_1|B) = P(e_1|h_1, B)P(h_1) + P(e_1|h_0, B)P(h_0)$$

So,

$$\frac{1}{P(h_1|e_1, B)} = 1 + \frac{P(e_1|h_0, B)P(h_0)}{P(e_1|h_1, B)P(h_1)}$$

Hence, an equivalent statement of (9.3)

$$1/P(h_1|b_0, e_1) > 1/P(h_1|b_1, e_1)$$

is equivalent to

$$1 + \frac{P(e_1|h_0, b_0)P(h_0)}{P(e_1|h_1, b_0)P(h_1)} > 1 + \frac{P(e_1|h_0, b_1)P(h_0)}{P(e_1|h_1, b_1)P(h_1)}$$

which simplifies to

$$\frac{P(e_1|h_0, b_0)}{P(e_1|h_1, b_0)} > \frac{P(e_1|h_0, b_1)}{P(e_1|h_1, b_1)}.$$

Finally, we rearrange this:

$$\frac{P(e_1|h_0, b_0)}{P(e_1|h_0, b_1)} > \frac{P(e_1|h_1, b_0)}{P(e_1|h_1, b_1)}. \tag{9.4}$$

So, the time constraint has more dramatic effect on the people that are not health-conscious. Health-conscious people are less likely to exercise if they are busy, but this probability drop is much less severe because they try to make time for exercise.

Excercise 9.1. *Give a realistic example where* (9.4) *fails to hold.*

Answer. In the same graphical structure as H, E, B but replace H with horrible weather, B with road block (due to construction) and E with extremely long commute.

Suppose the road can't handle traffic given either of the bad conditions. The difference between $P(e_1|h_0, b_0)$ and $P(e_1|h_0, b_1)$ is big, making the LHS small. But if the road is already hit with horrible weather, then the additional effect of road block is small, so the RHS is big.