Bayes Rule

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Which is shorthand for:

$$P(Y = y_i|X = x_j) = \frac{P(X = x_j|Y = y_i)P(Y = y_i)}{P(X = x_j)}$$

Bayes' Rule

- Product rule $P(a \wedge b) = P(a \mid b) P(b) = P(b \mid a) P(a)$
  $\Rightarrow$ Bayes' rule: $P(a \mid b) = P(b \mid a) P(a) / P(b)$

- or in distribution form

$$\mathbf{P}(Y|X) = \mathbf{P}(X|Y)\,\mathbf{P}(Y) / \mathbf{P}(X) = \alpha\mathbf{P}(X|Y)\,\mathbf{P}(Y)$$

- Useful for assessing diagnostic probability from causal probability:
  - P(Cause|Effect) = P(Effect|Cause) P(Cause) / P(Effect)

  - E.g., let *M* be meningitis, *S* be stiff neck:
    $P(m|s) = P(s|m) P(m) / P(s) = 0.5 \times 0.0002 / 0.05 = 0.0002$
  - Note: posterior probability of meningitis still very small!

Bayes' Rule and conditional independence

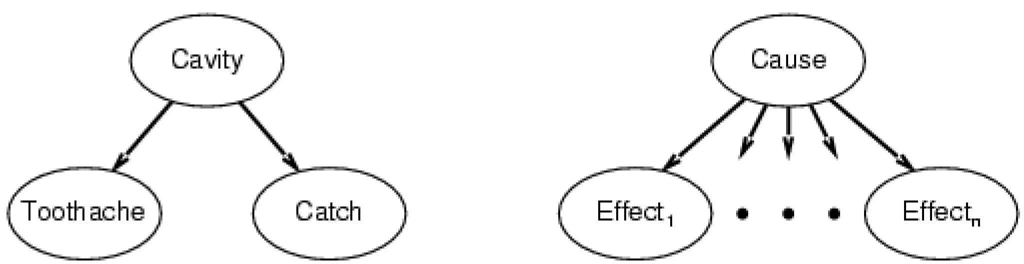
$\mathbf{P}(Cavity \mid toothache \wedge catch)$

$= \alpha \cdot \mathbf{P}(toothache \wedge catch \mid Cavity)\ \mathbf{P}(Cavity)$

$= \alpha \cdot \mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(catch \mid Cavity)\ \mathbf{P}(Cavity)$

- This is an example of a naïve Bayes model:

  $\mathbf{P}(\text{Cause},\text{Effect}_1, \ldots ,\text{Effect}_n) = \mathbf{P}(\text{Cause})\ \pi_i \mathbf{P}(\text{Effect}_i|\text{Cause})$



- Total number of parameters is linear in *n*

Naïve Bayes Classifier

- Calculate most probable function value

$V_{map} = \text{argmax } P(v_j | a_1, a_2, \ldots, a_n)$

$$= \text{argmax } \frac{P(a_1, a_2, \ldots, a_n | v_j) P(v_j)}{P(a_1, a_2, \ldots, a_n)}$$

$$= \text{argmax } P(a_1, a_2, \ldots, a_n | v_j) P(v_j)$$

Naïve assumption: $P(a_1, a_2, \ldots, a_n) = P(a_1)P(a_2) \ldots P(a_n)$

Naïve Bayes Algorithm

NaïveBayesLearn(*examples*)

For each target value v_j

$P'(v_j) \leftarrow$ estimate $P(v_j)$

For each attribute value a_i of each attribute a

$P'(a_i|v_j) \leftarrow$ estimate $P(a_i|v_j)$

ClassfyingNewInstance(x)

$$v_{nb} = \underset{v_j \in V}{\text{argmax}}\ P'(v_j) \prod_{a_j \in x} P'(a_i|v_j)$$

An Example

(due to MIT's open coursework slides)

f_1	f_2	f_3	f_4	**y**
0	1	1	0	**1**
0	0	1	1	**1**
1	0	1	0	**1**
0	0	1	1	**1**
0	0	0	0	**1**
1	0	0	1	**0**
1	1	0	1	**0**
1	0	0	0	**0**
1	1	0	1	**0**
1	0	1	1	**0**

R_1(1,1) = 1/5: fraction of all positive examples that have feature 1 = 1
R_1(0,1) = 4/5: fraction of all positive examples that have feature 1 = 0

R_1(1,0) = 5/5: fraction of all negative examples that have feature 1 = 1
R_1(0,0) = 0/5: fraction of all negative examples that have feature 1 = 0

Continue calculation of R_2(1,0) …

An Example

(due to MIT's open coursework slides)

f_1	f_2	f_3	f_4	**y**
0	1	1	0	**1**
0	0	1	1	**1**
1	0	1	0	**1**
0	0	1	1	**1**
0	0	0	0	**1**
1	0	0	1	**0**
1	1	0	1	**0**
1	0	0	0	**0**
1	1	0	1	**0**
1	0	1	1	**0**

(1,1) (0,1) (1,0) (0,0)

$\mathbf{R_1}$ = 1/5, 4/5, 5/5, 0/5

$\mathbf{R_2}$ = 1/5, 4/5, 2/5, 3/5

$\mathbf{R_3}$ = 4/5, 1/5, 1/5, 4/5

$\mathbf{R_4}$ = 2/5, 3/5, 4/5, 1/5

New x = <0, 0, 1, 1>

$S(1) = R_1(0,1)*R_2(0,1)*R_3(1,1)*R_4(1,1) = .205$

$S(0) = R_1(0,0)*R_2(0,0)*R_3(1,0)*R_4(1,0) = 0$

S(1) > S(0), so predict v = 1.