Probability

$P(X)$ – the probability of ~~not~~ event X occurring

$P(\text{Heads}) = 0.5$

$P(\text{Twins}) = 0.03$

$P(X|Y)$ – prob of event X given that Y occurs

$P(A|4.0) = 0.95$

$P(A|\neg 4.0) = 0.7$

$P(\text{Snow in 10 min} | \text{Snowing}) = 0.8$

$P(\text{Heads} | \text{just got heads}) = 0.5$

$P(x)P(y)$ = prob that both x and y occur

$P(H \text{ 1st time})\, P(H \text{ 2nd time}) = 0.5(0.5) = 0.25$

Bayes theorem

$$P(x|y) = \frac{P(y|x)}{P(y|x) + P(y|\neg x)}$$

example: $P(S|\text{"rich"}) = \dfrac{P(\text{"rich"}|S)}{P(\text{"rich"}|S) + P(\text{"rich"}|\neg S)}$

% of spam emails that have "rich" in them in the training dataset

$$\frac{0.7}{0.7 + 0.1} = \frac{0.7}{0.8} = 90\%$$

"the" $\dfrac{.9}{.9 + .9} = 0.5$

Suppose we have words $w_1, \ldots, w_n$.

X_i - event that a message contains word w_i.

Assume that we know $P(X_i|S)$ and $P(X_i|\neg S)$ (computed from training data)

Assume that X_i and X_j are independent — e.g. whether "rich" appears is independent of whether "rogaine" appears

reason its called "Naive" Bayes classifier

$$P(S|X=x) = \frac{P(X=x|S)}{P(X=x|S) + P(X=x|\neg S)}$$

Certain words appear

computed $P(X=x|S) = \prod P(X_i|S)$

example R = "rich"

M = "money"

$P(R \text{ and } M|S) = (0.2)(0.3) = 0.06$

$P(R \text{ and } M|\neg S) = 0.01(0.02) = 0.0002$

$$P(S|R \text{ and } M) = \frac{0.06}{0.06 + 0.0002} \approx 1$$