BAYESIAN CLASSIFICATION

- ***Probabilistic learning:*** Calculate explicit probabilities for hypothesis, among the most practical approaches to certain types of learning problems

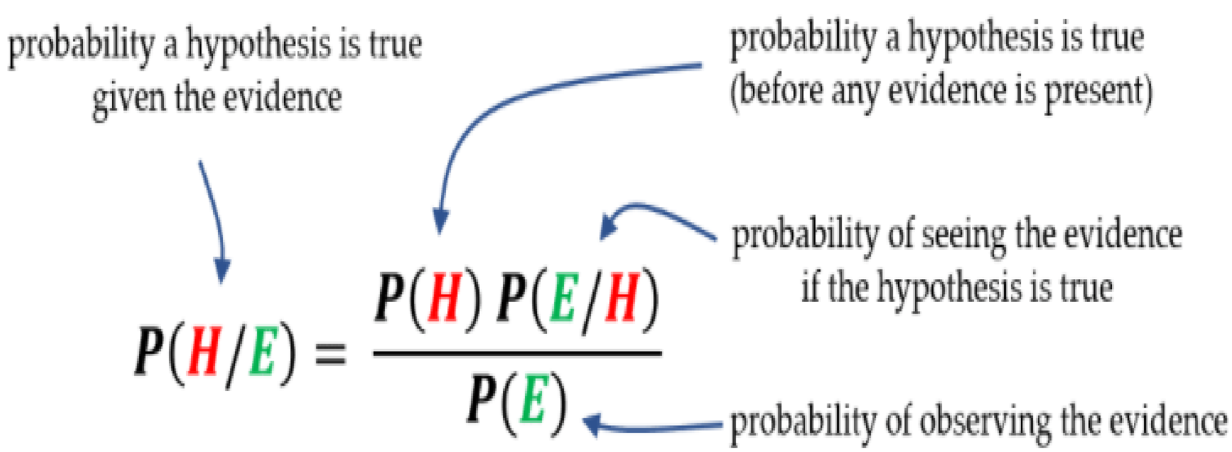
• ***Incremental:*** Each training example can incrementally increase/decrease the probability that a hypothesis is correct. Prior knowledge can be combined with observed data.

• ***Probabilistic prediction:*** Predict multiple hypotheses, weighted by their probabilities

• ***Standard:*** Even when Bayesian methods are computationally intractable, they can provide a standard of optimal decision making against which other methods can be measured

Given training data D, posteriori probability of a hypothesis h, P(h|D) follows the Bayes theorem

The formula for Bayes' theorem is



MAP (maximum posteriori) hypothesis

$$
\begin{aligned}
h_{map} &= \arg\max_{h \in H}(P(h \mid D)) \\
&= \arg\max_{h \in H}\left(\frac{P(D \mid h)P(h)}{P(D)}\right) \\
&= \arg\max_{h \in H}(P(D \mid h)P(h))
\end{aligned}
$$

Practical difficulty: require initial knowledge of many probabilities, significant computational cost

Naïve Bayes Classifier (I)

- A simplified assumption: attributes are conditionally independent:
- Greatly reduces the computation cost, only count the class distribution.

Naive Bayesian Classifier (II)

Given a training set, we can compute the probabilities

BAYESIAN CLASSIFICATION Algorithm

The classification problem may be formalized using a-posteriori probabilities:

- P(C|X) = prob. that the sample tuple
- X=<x1,…,xk> is of class C.
- E.g. P(class=N | outlook=sunny, windy=true,…)
- Idea: assign to sample X the class label C such that P(C|X) is maximal

Estimating a-posteriori probabilities

Bayes theorem:

$P(C|X) = P(X|C)\cdot P(C) / P(X)$

- $P(X)$ is constant for all classes
- $P(C)$ = relative freq of class C samples
- C such that $P(C|X)$ is maximum = C such that $P(X|C)\cdot P(C)$ is maximum
- Problem: computing $P(X|C)$ is unfeasible!