Naive Bayes Classifier

Along with decision trees, neural networks, nearest nbr, one of the most practical learning methods.

When to use

• Moderate or large training set available
• Attributes that describe instances are conditionally independent given classification

Successful applications:

• Diagnosis
• Classifying text documents
Naive Bayes Classifier

Assume target function $f : X \to V$, where each instance $x$ described by attributes $\langle a_1, a_2 \ldots a_n \rangle$. Most probable value of $f(x)$ is:

$$v_{MAP} = \arg\max_{v_j \in V} P(v_j | a_1, a_2 \ldots a_n)$$

$$v_{MAP} = \arg\max_{v_j \in V} \frac{P(a_1, a_2 \ldots a_n | v_j) P(v_j)}{P(a_1, a_2 \ldots a_n)}$$

$$= \arg\max_{v_j \in V} P(a_1, a_2 \ldots a_n | v_j) P(v_j)$$

Naive Bayes assumption:

$$P(a_1, a_2 \ldots a_n | v_j) = \prod_i P(a_i | v_j)$$

which gives

**Naive Bayes classifier:** $v_{NB} = \arg\max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j)$
Naive Bayes Algorithm

Naive\_Bayes\_Learn(examples)

For each target value \( v_j \)

\[ \hat{P}(v_j) \leftarrow \text{estimate } P(v_j) \]

For each attribute value \( a_i \) of each attribute \( a \)

\[ \hat{P}(a_i|v_j) \leftarrow \text{estimate } P(a_i|v_j) \]

Classify\_New\_Instance(\( x \))

\[ v_{NB} = \arg\max_{v_j \in V} \hat{P}(v_j) \prod_{a_i \in x} \hat{P}(a_i|v_j) \]
Naive Bayes: Example

Consider *PlayTennis* again, and new instance

\( \langle Outlk = \text{sun}, Temp = \text{cool}, \text{Humid} = \text{high}, Wind = \text{strong} \rangle \)

Want to compute:

\[
v_{NB} = \text{argmax} \ P(v_j) \prod_{i} P(a_i|v_j)
\]

\[
\begin{align*}
P(y) & \ P(\text{sun}|y) \ P(\text{cool}|y) \ P(\text{high}|y) \ P(\text{strong}|y) = .005 \\
P(n) & \ P(\text{sun}|n) \ P(\text{cool}|n) \ P(\text{high}|n) \ P(\text{strong}|n) = .021
\end{align*}
\]

\[\rightarrow v_{NB} = n\]