Directed Graphical Models

Daniel Khashabi[1]

KHASHAB2@ILLINOIS.EDU

0.1 Naive Bayes

In *Naive Bayes Classifier*, we assume that independence of input features, given the class, that is, we can write their joint distribution in the following way:

$$f(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n | G = g) = \prod_{i=1}^{n} f(\mathbf{X}_i | G = g)$$

Note that in the above formulation, $\mathbf{X}_i$ refer to each of input features, each of which could be of several dimension(and should not be confused with dimensions of one single feature, though in some cases they can be tough as the same thing). By applying this assumption in equation **??**, we have:

$$
\begin{aligned}
\log \frac{\Pr(G = k | X = x)}{\Pr(G = l | X = x)} &= \log \frac{\pi_l f(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n | G = l)}{\pi_k f(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n | G = k)} \\
&= \log \frac{\pi_l \displaystyle\prod_{i=1}^{n} f(\mathbf{X}_i | G = l)}{\pi_k \displaystyle\prod_{i=1}^{n} f(\mathbf{X}_i | G = k)} \\
&= \log \frac{\pi_l}{\pi_k} + \sum_{i=1}^{n} \log \frac{f(\mathbf{X}_i | G = l)}{f(\mathbf{X}_i | G = k)}
\end{aligned}
$$