Naive Bayes' classifier prediction step

7

$$
\begin{aligned}
\hat{\mathbf{y}} &= \underset{y}{\operatorname{argmax}} \; \log P(\mathbf{x}, y; \boldsymbol{\mu}, \boldsymbol{\phi}) \\
&= \underset{y}{\operatorname{argmax}} \; \log P(\mathbf{x} \mid y; \boldsymbol{\mu}, \boldsymbol{\phi}) + \log P(y; \boldsymbol{\mu}) \\
&= \underset{y}{\operatorname{argmax}} \; \log \left[B(\mathbf{x}) \prod_{j=1}^{N} \phi_j^{\mathbf{x}_j} \right] + \log \mu_y \\
&= \underset{y}{\operatorname{argmax}} \; \log B(\mathbf{x}) + \sum_{j=1}^{N} x_j \log \phi_{y,j} + \log \mu_y \\
&= \underset{y}{\operatorname{argmax}} \; \sum_{j=1}^{N} x_j \log \phi_{y,j} + \log \mu_y
\end{aligned}
$$

This ends up being the label with the highest log posterior probability

Naive Bayes' classifier prediction step

$$
\begin{aligned}
\hat{\mathbf{y}} &= \underset{y}{\operatorname{argmax}}\ \log P(\mathbf{x}, y; \boldsymbol{\mu}, \boldsymbol{\phi}) \\
&= \underset{y}{\operatorname{argmax}}\ \log P(\mathbf{x} \mid y; \boldsymbol{\mu}, \boldsymbol{\phi}) + \log P(y; \boldsymbol{\mu}) \\
&= \underset{y}{\operatorname{argmax}}\ \log \left[B(\mathbf{x}) \prod_{j=1}^{N} \phi_j^{\mathbf{x}_j} \right] + \log \mu_y \\
&= \underset{y}{\operatorname{argmax}}\ \log B(\mathbf{x}) + \sum_{j=1}^{N} x_j \log \phi_{y,j} + \log \mu_y \\
&= \underset{y}{\operatorname{argmax}}\ \sum_{j=1}^{N} x_j \log \phi_{y,j} + \log \mu_y
\end{aligned}
$$

This ends up being the label
with the highest log posterior
probability

Naive Bayes' classifier estimation step

Maximum likelihood estimate of ϕ_y

$$\hat{\phi}_{y,j} = \frac{\sum_{i\,:\,y^{(i)}=y}^{M} x_j^{(i)}}{\sum_{j'=1}^{N} \sum_{i\,:\,y^{(i)}=y}^{M} x_{j'}}$$

i.e., the ratio between the number of times label *y* appears in conjunction with word x_j, and the number of times label *y* appears in total.