Naive Bayes' classifier

- Assumption 1: The text and label in one document does not affect those of another.

- Assumption 2: Words in a sentence are independent, conditioned on the class label

- Taken from Eisenstein, 2019, Chp 2

Algorithm 1 Generative process for the Naïve Bayes classification model

for Instance $i \in \{1, 2, \ldots, M\}$ **do**:
  Draw the label $y^{(i)} \sim \text{Categorical}(\boldsymbol{\mu})$;
  Draw the word counts $\boldsymbol{x}^{(i)} \mid y^{(i)} \sim \text{Multinomial}(\boldsymbol{\phi}_{y^{(i)}})$.

chain rule $P(\mathbf{x}, y) = P(\mathbf{x} \mid y)P(y)$

where $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_K]$ label probability

$\boldsymbol{\phi} = [\phi_1, \ldots, \phi_N]$ word probability

$$P_{mult}(\mathbf{x} \mid y; \phi) = B(\mathbf{x}) \prod_{j=1}^{N} \phi_j^{\mathbf{x}_j} \qquad B(\mathbf{x}) = \frac{\left(\sum_{j=1}^{N} \mathbf{x}_j\right)!}{\prod_{j=1}^{N} \mathbf{x}_j!}$$

Naive Bayes' classifier prediction step

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