Lecture 17 (Naive Bayes)

Note: $y = b$ anywhere in this lecture is for $b = \{0, 1\}$

It is a generative model

1 Assumptions

1. $y \sim \text{Bernoulli}(\phi)$ (y is usually considered to be discrete)
2. x is discrete taking values $1, 2, \ldots, L$
3. $x_i \perp x_j | y$, that is, all x_i are independent given y
4. (repeat) **Only the conditional probability is independent**
5. This is the only assumption made by naive Bayes

2 Mathematical Analysis

$$P(x|y) = \prod_{i=1}^{n} P(x_i|y)$$

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} = \frac{\prod_{i=1}^{n} P(x_i|y)}{\sum_y \prod_{i=1}^{n} P(x_i|y)}$$

Now, to find the class, we can directly compute:

$$\underset{y}{argmax} P(y|x) = \underset{y}{argmax} P(x|y)P(y)$$

3 Generating the Model

$$x_{j|y=b} = \text{Multinoulli}(\theta_{j|y=b})$$

$$\theta_{j|y=b} = \left(\theta_{j1|y=b}, \ldots, \theta_{jL|y=b}\right)$$

In the above expression, $\displaystyle\sum_{l=1}^{L} \theta_{jl|y=b} = 1$

Now we compute *argmax* for $LL(\Theta) = \log \prod_{i=1}^{m} P(x_i, y_i; \Theta)$,

$$LL(\Theta) = \sum_{i=1}^{m} (\log P(y_i; \phi) + \log P(x_i|y_i; \Theta))$$

The second term can be written as:

$$\sum_{j=1}^{n} \log P((x_i)_j | y; \Theta)$$

Now the above is *simplified* by adding probabilities for $y = 1$ and $y = 0$ and writing $P((x_i)_j | y; \Theta)$ and product of probabilities for each component.

(The solving has been skipped for sanity purposes)

On computing $\nabla_{\Phi} LL(\theta) = 0$, we get:

$$\phi = \frac{\sum_{i=1}^{m} y_i}{m}$$

$$\theta_{jl|y=b} = \frac{\sum_{i=1}^{m} (y_i = b)(x_j = l)}{\sum_{i=1}^{m} y_i = b}$$

3.1 Gaussian Naive Bayes Model

The assumptions made are that x_j is independent and $x_j | y = b$ follows normal distribution. This gives Σ as a diagonal matrix of σ_{jb}.