COL774

Machine Learning

Sep 10, 2021

Last class:- Naive Bays Model

$x_1 \ldots x_n$ | y :- target variable

$$P(x, y) = P(y) \prod_{j=1}^{n} P(x_j | y)$$

$$\log\left[P(x^{(i)}, y^{(i)}; \theta)\right] = \log\left(\prod_{i=1}^{m} P(y^{(i)}; \theta) \prod_{j=1}^{n} P(x_j^{(i)} | y^{(i)}; \theta)\right)$$

↳ $x_j^{(i)} \sim$ Multinoulli $(\theta_j | y=1)$

$(\theta_j | y=0)$

↳ $x_j^{(i)} \sim$ Normal (μ_j, σ_j^2)

↳ Gaussian Naive Bays

→ class prior

$$\phi = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}{m}$$

prior

Π

$$\theta_{jl} | y=1 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\} \, 1\{x_j^{(i)} = l\} + 1 \alpha}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\} + L \cdot \alpha}$$

$$\theta_{jl} | y=0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} \, 1\{x_j^{(i)} = l\} + 1 \alpha}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} + L \alpha}$$

Laplace Smoothing

Smoothing of the parameters → $\boxed{\text{MAP estimate}}$

↓

prior

Flu :- y

$\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$ Cough Cold Fever Diarrohea

$\in \{0, 1\}$

$\cancel{\text{Cough}} = 0 \mid \text{Flu} = 1$ if $x_j = 1$

$\theta_{j\ell|y=1}$

$$\Rightarrow \theta_{j\ell|y=1} = 0$$

$\ell = 0$, j : cough

$$\Rightarrow p(y=1|x) = ? \quad \frac{\prod_{j=1}^{n} p(x_j|y=1)\, p(y=1)}{p(x)}$$

Flu; x_j :- cough $= 0$

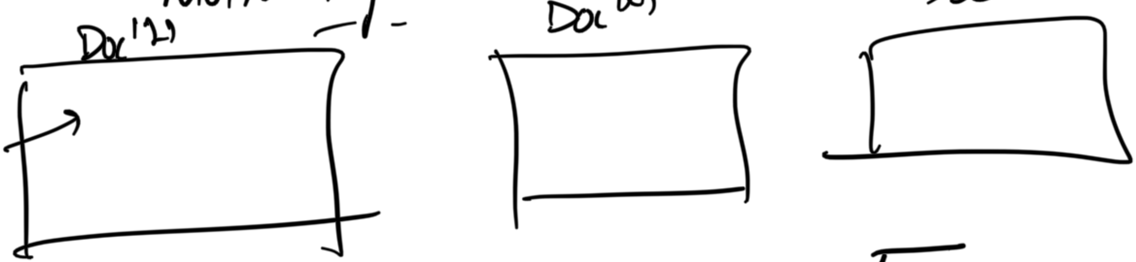
$= 0$

Fever: high
Cold: Yes
Diarrohea: Yes

Text Classification:

Motivating Example:

Doc$^{(1)}$ Doc$^{(i)}$ Doc$^{(m)}$



$y^{(i)}$: $\{1 \ldots r\}$

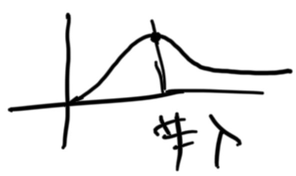
$k \in \{1 \ldots r\}$

$\Sigma \; \{w_1 \ldots w_{|V|}\} \to$ vocabulary $\to$ table

$\to$ Dictionary

$\Downarrow$ Naive Bayes

Classification $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$



$\# \lambda$

What could be features?

$x_j^{(i)}$:- Frequencies of words in document

$x_j^{(i)} | y=k \sim \text{Poisson}(\lambda_j | y=k)$]

Bernoulli Event Model

$y^{(i)} \sim \text{Multinoulli}(\Phi)$

$\Phi = (\Phi_1 \ldots \Phi_r) \quad \sum_k \Phi_k = 1$

$x_j^{(i)} \in \{0, 1\}$

0 → w_j :- is not present in the document

1 → w_j is present in the document

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

$\underbrace{\theta_{jl|y=k}}_{P(x_j^{(i)} = l \mid y=k)}$ $\quad l \in \{0, 1\}$

Limiting :- This does not model counts.

Multinomial Event Model

$\boxed{\{w_1 \ldots w_{|V|}\} \quad \theta_{jl|y=k}}$

$x_j^{(i)}$:- ∀ every position in the document.

$n^{(i)}$:- # words in ith document

$j \in \{1 \ldots n^{(i)}\}$

$x_j^{(i)} \in \{1 \ldots |V|\}$

⇓

$x_j^{(i)} \mid y=k \sim \text{Multinoulli}(\theta_{jl|y=k})$

$l \in \{1 \ldots |V|\}$

Generation :-

→ $y^{(i)} \sim \text{Multinoulli}(\Phi)$

→ $y^{(i)} = k$

$x_j^{(i)} \mid y=k \sim \text{Multinoulli}(\theta_{jl|y=k})$

$\forall j : j \in \{1 \ldots n^{(i)}\}$

$$\sum_{l=1}^{|V|} \theta_{jl|y=k} = 1$$

⇒ Different features for different documents

parameters of the model :-

parameters of the model:

$$O\left(|V| \cdot \max_i (n^{(i)}) \cdot \gamma\right)$$

~~Eq~~ Another simplifying assumption:

["Bag of words"]

$$\theta_{j\ell|y=k} = \theta_{j'\ell|y=k} \quad \forall j, j'$$

Parameter tying assumption

Position invariant :- $O\left(|V| \cdot \gamma\right)$

(HW) Naive Bayes model:- ML estimate $\{x^{(i)}, y^{(i)}\}_{i=1}^{m}$

Φ :- $(\phi_1 \ldots \phi_\gamma)$

$$\phi_k = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = k\}}{m}$$

$$\forall j :- \theta_{j\ell|y=k} = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = k\} \sum_{j=1}^{n^{(i)}} 1\{x_j^{(i)} = \ell\}}{\sum_{i=1}^{m} 1\{y^{(i)} = k\} \sum_{j=1}^{n^{(i)}}}$$

$$= \frac{\left[\sum_{i=1}^{m} 1\{y^{(i)} = k\}\, n_\ell^{(i)}\right] + 1}{\left[\sum_{i=1}^{m} 1\{y^{(i)} = k\}\, n^{(i)}\right] + |V|\cdot k}$$

$n_\ell^{(i)} \Rightarrow$ # of times we 'l' occurs in $x^{(i)}$

Smoothed estimate

$\Rightarrow$ $\left[\begin{array}{l} \cancel{MAP} :- ML \\ \underset{\theta}{\operatorname{argmax}}\, P(D;\theta) \\ \underset{\theta}{\operatorname{argmax}}\, P(D|\theta) \end{array}\right]$

MAP

$\left[\underset{\theta}{\operatorname{argmax}}\, P(\theta|D)\right] \Downarrow$

$P(\theta)$:-

Bayesian Estimation