BPE

aaabcaabdeab

① $A = aa \quad B = ab$

ABcAbdeB

② vocab $= \{A, B, b, c, d, e\}$

NB inference

$$y = \underset{y \in Y}{\arg\max} \frac{P_\theta(x|y)\,\cancel{P(y)}^{\,1/2}}{\cancel{P(x)}}$$

$$= \underset{y \in Y}{\arg\max} \prod_{i=1}^{n} P_\theta(x_i|y)$$

$$= \underset{y \in Y}{\arg\max} \sum_{i=1}^{n} \underbrace{\log P_\theta(x_i|y)}_{\text{score of each word}}$$

MLE for NB

Likelihood function

$$L(\theta, \alpha) = \sum_{i=1}^{N} \log p(x^{(i)}, y^{(i)}; \theta, \alpha)$$

$$= \sum_{i=1}^{N} \log \underbrace{p(x^{(i)} | y^{(i)}; \theta)}_{\text{const.}} \, p(y^{(i)}; \alpha)$$

$$\theta \in \mathbb{R}^{2 \times |V|} \quad \text{s.t. } \sum_i \theta_{i,y} = 1$$

$$\alpha \in \mathbb{R}$$

$$p(y^{(i)}; \alpha) = \begin{cases} \alpha & \text{if } y^{(i)} = 1 \\ 1-\alpha & \text{o.w.} \end{cases}$$

$$= \mathbb{1}(y^{(i)}=1)\,\alpha + \mathbb{1}(y^{(i)}=0)(1-\alpha)$$

$L(\alpha)$ is concave.

$$\frac{\partial L(\alpha)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{N} \log\left[\mathbb{1}(y^{(i)}=1)\,\alpha + \mathbb{1}(y^{(i)}=0)(1-\alpha)\right]$$

$$= \frac{\partial}{\partial \alpha} \sum_{i=1}^{N} \mathbb{1}(y^{(i)}=1) \log \alpha + \mathbb{1}(y^{(i)}=0) \log(1-\alpha)$$

$$= \sum_{i=1}^{N} \mathbb{1}(y^{(i)}=1) \frac{1}{\alpha} - \mathbb{1}(y^{(i)}=0) \frac{1}{1-\alpha} = 0$$

$$\hat{\alpha} = \sum_{i=1}^{N} \mathbb{1}(y^{(i)}=1) \Big/ \sum_{i=1}^{N} \mathbb{1}(y^{(i)}=1) + \mathbb{1}(y^{(i)}=0)$$

$$= \#\text{pos. ex.} / \#\text{total}$$

MLE for LR

$$L(w) = \sum_{i=1}^{N} \log P(y^{(i)} \mid x^{(i)}; w)$$

$$= \sum_{i=1}^{N} y^{(i)} \log \frac{1}{1+e^{\underbrace{-w\cdot\phi(x^{(i)})}_{z}}} + (1-y^{(i)}) \log\left[1 - \frac{1}{1+e^{-w\cdot\phi(x)}}\right]$$

Is $L(w)$ concave? $\log \frac{1}{1+e^{-z}}$

① plot
② $f''(z) \le 0$

No closed-form solution!

BoW

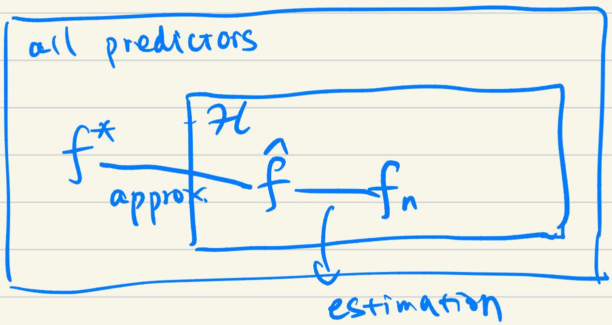
in for a penny, in for a pound

$$\text{in} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}_{|V|\times 1} \begin{array}{l} \to \mathbb{1}(\text{"in"} = w_1) \\ \vdots \\ \to \mathbb{1}(\text{"in"} = w_{20}) \\ \\ \to \mathbb{1}(\text{"in"} = w_{|V|}) \end{array}$$

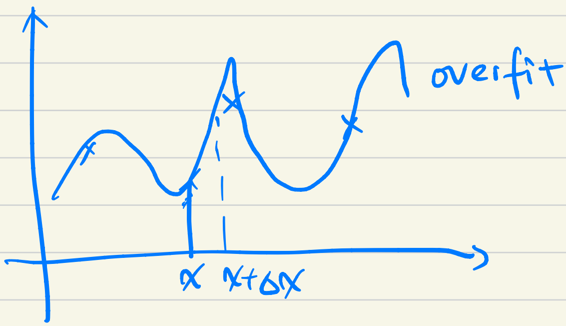
$$\phi_{BoW}(\text{sentence}) = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} \begin{array}{l} \text{the} \\ \text{a} \\ \text{an} \\ \text{in} \\ \text{for} \\ \text{penny} \\ \text{pound} \\ \vdots \end{array}$$

$|V|\times 1$

Error decomposition



Why small norms?



$$|w \cdot x - w \cdot (x + \Delta x)|$$

$$= |w \cdot \Delta x|$$

$$\leq \underbrace{\|w\|}_{\text{small}} \cdot \|\Delta x\|$$