

Mathematical details

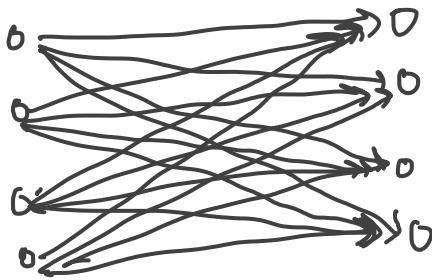
20 June 2023 09:06

Bigram LM

Assumption

$$P(y^{(1)}, y^{(2)}, \dots, y^{(T)}) \\ = P(y^{(1)}) P(y^{(2)} | y^{(1)}) \dots \\ \dots P(y^{(T)} | y^{(T-1)})$$

Training strategy



$$\underline{x} = (\text{vocab.size}, 1) \quad \underline{y} = (\text{vocab.size}, 1)$$

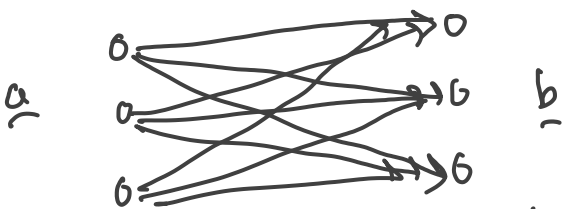
lets say you have

$$\text{vocab} = [a, b, c]$$

Now if want to get

$$P(b|a) \text{ where } \text{one-hot}(a) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{one-hot}(b) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$



W = weight matrix can be defined

$$-\log \begin{pmatrix} P(a) & P(b|a) & P(c|a) \\ P(b) & P(a|b) & P(c|b) \\ P(c) & P(a|c) & P(b|c) \end{pmatrix}$$