Machine learning.

Non-Coding Portion.

4a. PROOF of two equations

$\mathcal{S}$ → bag of words (document)

$n(w)$ → each word $w \in W$ appears in $\mathcal{S}$.

each word independent

Assume that the document was generated one word at a time.

~~it does not matter.~~

Assume that words are drawn from the same distribution $\theta = \theta_w$ where $w \in W$ appears w probability θ_w.

No it does not matter. Since, as long as the word, in the dictionary, w do not appear it is counted as $n(w) = 0$.

b. $\frac{\partial \ell}{\partial \mu} = 0$ } to derive μ

Training Loss. → $$\mathcal{L}_n(\mu, \sigma^2) = \frac{d}{2}\log(2\pi\sigma^2) + \frac{1}{2n\sigma^2}\sum_{i=1}^{n}{}_{x \in S}\|x-\mu\|^2$$

$$0 = 0 + \frac{\cancel{2}}{\cancel{2}n\sigma}\sum_{i=1}^{n}{}_{x \in S}\|x-\mu\|$$

$$0 = \sum_{i=1, x \in S}^{n} x - n\mu$$

$$\mu = \frac{1}{n}\sum_{i=1, x \in S}^{n} x_i \quad \#$$

$\frac{\partial \ell}{\partial \sigma} = 0$ } to derive σ $\log_e = \ln$

$$0 = \frac{d}{\cancel{2}}\left(\frac{\cancel{4\pi}\sigma}{\cancel{2\pi}\sigma^2}\right) + \frac{-\cancel{2}}{\cancel{2}n\sigma^3}\sum_{i=1}^{n}{}_{x \in S}\|x-\mu\|^2$$

$$\frac{1}{nd}\sum_{i=1}^{n}{}_{x \in S}\|x-\mu\|^2 = \sigma^2$$

$$\sigma = \sqrt{\frac{1}{nd}\sum_{i=1}^{n}\|x-\mu\|^2} \quad \#$$

AZONE

Scanned by CamScanner