Lecture 9 Naive Bayes, MLE

$P(x|y)$ $\qquad$ $P(y)$

$$\underline{P(y|x)} \propto P(x|y)\,P(y)$$

1. count of the words

2. order

$P(y)$ <u>Bernou</u>

y spam or not spam

? $\quad$ 10

x

size of (dict)

50K

$$\boxed{2^{50K}}$$

y

$$P(x_i \mid y_j) = P(x_i \mid y, x_j)$$

computer product

electrical

sell

price

$$P(x_1, x_2, x_3 \cdots x_n \mid C_1\ C_2\ C_3)$$

$$= P(x_1 \mid C_1 \cdots C_2)\, P(x_2 \mid x_1, C_1\ C_2 \cdots C_3)\, P(x_3 \mid x_1, x_2) \cdots$$

$$\cdots P(x_n \mid x_1 \cdots x_{n-1})$$

$$= P(x_5)\, P(x_2 \mid x_5)\, P(x_3 \mid x_2\, x_5)$$

$$P(x_n \mid x_3\ x_2\ x_5) \cdots$$

$$P(x_1 \mid y, \cdots x_2 \cdots x_3 \cdots)$$

$$= P(x_1 \mid y)$$

$$2^{50K}$$

$\mathbb{1}(\text{true}) = 1$

$\mathbb{1}(\text{false}) = 0$

$\wedge$ and

$P(\text{"computer" is present} \mid \text{spam})$

$=$

$\phi_y =$

normalization

$$\sum_{j=1}^{k} \phi_j = 1$$

$$\log P_\theta(x) + l(\theta)$$

$P_{data}(x)$ no parameter

$$\underbrace{E_{x\sim p(x)}\, f(x)}$$

$$\underbrace{E\left[\frac{1}{n}\sum_{i=1}^{n} f(x^{(i)})\right]} = \underbrace{E_{x\sim p(x)}\, f(x)}$$

what if $\underbrace{Var\left[\frac{1}{n}\sum_{i=1}^{n} f(x^{(i)})\right] = 1}$

$$n \longrightarrow \infty$$

$$Var = \frac{Var[f(x)]}{n}$$

$$\lim_{n\to\infty} Var = 0$$

$\boxed{E_{x\sim p(x)} f(x)}$

$p(y)$

$\downarrow$

y animal type

$\downarrow$ $p(x|y)$

x cat image

dog image

$p(y)$

$\downarrow$

y animal typ

$\downarrow$ $p(z|y)$ mutin… of

z action

$\downarrow$ $p(x|y,z)$

x Gaussian

$p(x)$

$$p(x) = \sum_{y,z} p(y,z,x)$$

$$= \sum_{y,z} p(y)\, p(z|y)\, p(x|y,z)$$

50K

x_0

ancestral sampling

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} (x-\mu)^2$$

inverse CDF

$$CDF = \int_x P(x)\, dx$$

$P(x)$

$$bias = \frac{1}{n}$$

$$\text{mean}(x) = \left(\frac{\sum_{i}^{n} x_i}{n}\right) \hat{\mu}$$

$\text{Var}(x)$ → biased

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2\right] = \text{Var}(x)$$

unbiased

$\text{Var}(x)$

$= E[x-\mu)^2$

central limit theorem

bias or unbias?

$$E\left[\frac{\sum_{i}^{n} x_i}{n}\right] = \text{mean}(x)$$

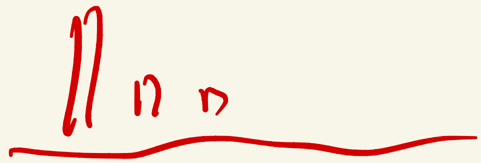
μ, σ^2

random variable

$\underbrace{P(\mu)}_{N(0,1)}$ $\quad P(\sigma^2)$

$$\underbrace{P(\text{course} = \text{CSE}}_{\text{history}}) = \underbrace{\theta}_{d}$$

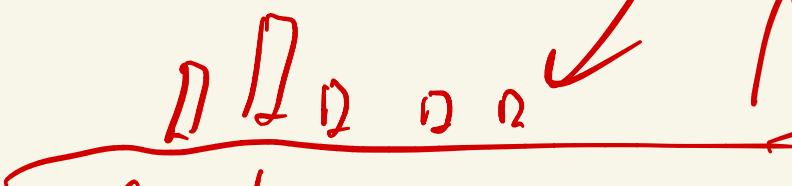
art



CSE

$$\underbrace{P(\theta)}$$

$\boxed{\text{Dirichlet}}$



CSE history

$P(\theta)$ $P(D|\theta)$

$P(\theta|D) \propto P(D|\theta) P(\theta)$

Bernoulli

Bern

$P(\theta)$

$P(\theta)$

θ

multinomial

discrete