HW-0 Representation Learning

(3) ~~Gaussian~~

Let N data samples $\{\vec{x}^1, \vec{x}^2, \vec{x}^3, \ldots \vec{x}^N\}$

& $\vec{x}^{(i)}$ is independent of $\vec{x}^{(j)}$

& $f(\vec{x}, \theta)$ be the model from which these datapoints are extracted from, θ being the model parameters

Finding θ is MLE problem

(d) Gaussian

Let $\vec{x}^{(i)} \sim N(\mu, \sigma)$

$$L(\vec{x}, \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\vec{x}^{(i)} - \mu)^2}{2\sigma^2}\right)$$

$$\log(L(\vec{x}, \mu, \sigma)) = \frac{-N}{2} \log(2\pi\sigma^2) + \sum_{i=1}^{N} \frac{-(\vec{x}^{(i)} - \mu)^2}{2\sigma^2}$$

Assuming univariate samples

& for Maximum Likelihood, $\frac{\partial}{\partial \mu} \log(L)$ & $\frac{\partial}{\partial \sigma} \log(L) = 0$

$$\frac{\partial}{\partial \mu}(\log(L(\vec{x}, \mu, \sigma))) = \cancel{2} \sum_{i=1}^{N} \left(\frac{\vec{x}^{(i)} - \mu}{\cancel{2\sigma^2}}\right) = 0$$

$$\Rightarrow \boxed{\mu_{ML} = \frac{1}{N} \sum_{i=1}^{N} \vec{x}^{(i)}}$$

$$\frac{\partial}{\partial \sigma}\left(L(x,\mu,\sigma)\right) = \frac{\partial}{\partial \sigma}\left[-\frac{N}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{N}\left(\frac{(x-\mu)^2}{2\sigma^2}\right)\right] = 0$$

$$\Rightarrow -\frac{N}{\cancel{2}}\left[\frac{1}{2\pi\sigma^2}\right]\left[\cancel{2}\sigma\right] + \left[\sum_{i=1}^{N}\left(\frac{(x-\mu)^2}{}\right)\right]\left[\frac{+1}{(2\sigma^2)^2}\right]\left[4\sigma\right] = 0$$

$$\frac{N}{2\pi\sigma} = \frac{1}{\sigma^2}\left[\sum_{i=1}^{N}\frac{(x-\mu)^2}{}\right]$$

$$\frac{N\sigma^2}{2\pi} = \sum_{i=1}^{N}(x-\mu)^2$$

$$\boxed{\sigma^2_{ML} = \frac{1}{N}\sum_{i=1}^{N}(x-\mu)^2}$$

(b) Poisson:-

$$f(x,\lambda) = \frac{\lambda^{x_i}e^{-\lambda}}{x!}$$

$$L(x,\theta) = \prod_{i=1}^{N}\frac{\lambda^{x_i}e^{-\lambda}}{x^{(i)}!}$$

$$\log L(x,\theta) = \sum_{i=1}^{N}\left[x_i\log\lambda - \lambda - \log(x_i!)\right]$$

$$= \sum_{i=1}^{N}x_i\log\lambda - N\lambda - \sum_{i=1}^{N}\left[\frac{(x_i)(x_i+1)}{2}\right]$$

$$\frac{\partial}{\partial\lambda}\log L(x,\theta) = 0 \text{ for } \lambda_{ML}$$

$$\Rightarrow \frac{1}{\lambda}\sum_{i=1}^{N}x_i - N = 0$$

$$\Rightarrow \boxed{\lambda_{ML} = \frac{1}{N}\sum_{i=1}^{N}x_i}$$

(c) Exponential

$$f(x_i,\lambda) = \begin{cases} \lambda e^{-\lambda x_i} & x_i \geq 0 \\ 0 & \text{else} \end{cases}$$

Lets say N observations of x_i are positive.

$$\Rightarrow \quad L(x,\lambda) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i}$$

$$\ln L(x,\lambda) = \sum_{i=1}^{N} \left[\ln \lambda - \lambda x_i \right]$$

$$\ln(L(x,\lambda)) = N \ln \lambda - \lambda \sum_{i=1}^{N} x_i$$

for λ_{ML}, $\frac{\partial}{\partial \lambda}\left(\ln(L(x,\lambda))\right) = 0$

$$\Rightarrow \quad \frac{N}{\lambda_{ML}} - \sum_{i=1}^{N} x_i = 0 \quad \Rightarrow \quad \boxed{\lambda_{ML} = \frac{N}{\sum_{i=1}^{N} x_i}}$$

(e) Laplacian

$$f(x,\mu,\beta) = \frac{1}{2\beta} e^{-\left\{\frac{|x-\mu|}{\beta}\right\}}$$

$$L(x,\mu,\beta) = \prod_{i=1}^{N} \frac{1}{2\beta} \exp\left\{ -\frac{|x_i-\mu|}{\beta} \right\}$$

$$\ln(L(x,\mu,\beta)) = -N \ln 2\beta - \sum_{i=1}^{N} \frac{|x_i-\mu|}{\beta}$$

for optimal solution

$$\frac{\partial}{\partial \beta} \ln(L) = 0 \quad \& \quad \frac{\partial}{\partial \mu} \ln(L) = 0.$$

$$\frac{\partial}{\partial \beta}(L) = 0$$

$$\Rightarrow \frac{\partial}{\partial \beta}\left[-N\ln(2\beta) - \sum_{i=1}^{N}\frac{|x_i-\mu|}{\beta}\right] = 0$$

$$\Rightarrow \left[\frac{-N}{\cancel{2}\beta}(\cancel{2}) + \frac{1}{\beta^2}\sum_{i=1}^{N}|x_i-\mu|\right] = 0$$

$$+N\beta = \sum_{i=1}^{N}|x_i-\mu|$$

$$\boxed{\beta_{ML} = \frac{\sum_{i=1}^{N}|x_i-\mu|}{N}}$$

$$\frac{\partial}{\partial \mu}(L) = 0$$

$$\Rightarrow \frac{\partial}{\partial \mu}\left[-N\ln(2\beta) - \sum_{i=1}^{N}\frac{|x_i-\mu|}{\beta}\right] = 0$$

$$\Rightarrow -\frac{1}{\cancel{\beta}}\sum_{i=1}^{N}\frac{|x_i-\mu|}{x_i-\mu} = 0$$

$$\Rightarrow \sum_{i=1}^{N}\operatorname{Sgn}(x_i-\mu) = 0$$

$\Rightarrow \sum \operatorname{Sgn}(x) = 0 \Rightarrow$ there are equal samples of x below zero and above zero

$\Rightarrow (x_i - \mu)$ has equal negatives and positives

$\Rightarrow \mu =$ middle value among the collection of data

$\Rightarrow \boxed{\mu = \text{median}(x_i)}$

(a) Binomial :-

$$p(x_i, \theta) = \theta^{x_i} (1-\theta)^{1-x_i} \quad \left\{ x_i \text{ is preferably an } \cancel{\text{integer}} \ \{0,1\} \right\}$$

$$L(x, \theta) = \prod_{i=1}^{N} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$\ln L(x, \theta) = \sum_{i=1}^{N} \left[x_i \ln\theta + (1-x_i)(\ln(1-\theta)) \right]$$

$$\frac{\partial}{\partial\theta} \left(\ln (L(x,\theta)) \right) = 0$$

$$\Rightarrow \frac{\partial}{\partial\theta} \left[\sum_{i=1}^{N} \left[x_i \ln\theta + (1-x_i) \ln(1-\theta) \right] \right] = 0$$

$$\Rightarrow \frac{1}{\theta} \sum_{i=1}^{N} x_i + \frac{-1}{1-\theta} \sum_{i=1}^{N} (1-x_i) = 0$$

$$\Rightarrow \left[\frac{1}{\theta} + \frac{1}{1-\theta} \right] \sum_{i=1}^{N} x_i = \frac{N}{1-\theta}$$

$$\theta \neq 1$$

$$\frac{1}{\theta \cancel{(1-\theta)}} \sum_{i=1}^{N} x_i = \frac{N}{\cancel{1-\theta}}$$

$$\boxed{\theta_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i}$$

If K trials:-

$$P(x_i, K, \theta) = {}^{K}C_{x_i} (\theta)^{x_i} (1-\theta)^{K-x_i}$$

$$L(x, N, \theta) = \prod_{i=1}^{N} {}^{K}C_{x_i} (\theta)^{x_i} (1-\theta)^{K-x_i}$$

$$\ln(L) = \sum_{i=1}^{N} \ln\left({}^{K}C_{x_i}\right) + x_i \ln\theta + (K-x_i)\ln(1-\theta)$$

$$\frac{\partial}{\partial\theta} \ln(L) = 0$$

$$\Rightarrow \frac{1}{\theta} \sum x_i + \frac{-1}{1-\theta} \sum K - x_i = 0$$

$$\Rightarrow \left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) \sum x_i = \frac{KN}{1-\theta}$$

$$\frac{\sum x_i}{\theta(\cancel{1-\theta})} = \frac{KN}{(\cancel{1-\theta})}$$

$$\boxed{\theta = \frac{1}{KN} \sum x_i}$$

$$\frac{\partial}{\partial K} \ln(L) = 0$$

$$\Rightarrow \frac{\partial}{\partial K}\left(\sum_{i=1}^{N} \left[\ln\left({}^{K}C_{x_i}\right)\right] + \sum_{i=1}^{N} K \ln(1-\theta)\right) = 0$$

$$\Rightarrow \frac{\partial}{\partial K}\left\{NK \ln(1-\theta)\right\} + \frac{\partial}{\partial K}\left\{\sum_{i=1}^{N} \frac{K!}{x!\,(K-x_i)!}\right\} = 0$$

$$N \ln(1-\theta) + \frac{\partial}{\partial K}\left\{\sum_{i=1}^{N} \ln \left(\frac{\Gamma(K)}{\Gamma(x)\,\Gamma(K-x)}\right)\right\}$$

$$\frac{\partial}{\partial k} \ln\left(\frac{\Gamma(k)}{\Gamma(x)\Gamma(k-x)}\right) = \frac{\Gamma(x)\Gamma(k-x)}{\Gamma(k)} \left[\frac{\{\Gamma(x)\Gamma(k-x)\}\Gamma(k)\left(-\gamma+\sum_{i=1}^{k}\frac{1}{i}\right) - \Gamma(k)\Gamma(x)\Gamma(k-x)\left\{-\gamma+\sum_{i=1}^{k-x}\frac{1}{i}\right\}}{\left(\Gamma(x)\Gamma(k-x)\right)^2}\right]$$

$$= \frac{\cancel{\Gamma(x)\Gamma(k-x)\Gamma(k)}}{\cancel{\Gamma(k)\Gamma(x)\Gamma(k-x)}}\left\{\cancel{-\gamma}+\cancel{\gamma}+\sum_{i=1}^{k}\frac{1}{i}-\sum_{i=1}^{k-x}\frac{1}{i}\right\} = \left\{\sum_{i=k-x+1}^{k}\frac{1}{i}\right\}.$$

$$\Rightarrow \frac{\partial}{\partial k}\sum_{i=1}^{N}\ln\left(\frac{\Gamma(k)}{\Gamma(x)\Gamma(k-x)}\right) = \left\{\sum_{i=1}^{N}\sum_{j=k-x_i+1}^{k}\frac{1}{i}\right\}.$$

$$\Rightarrow N\ln(1-\theta) + \sum_{i=1}^{N}\sum_{j=k-x_i+1}^{k}\frac{1}{i} = 0$$

where K is the reqd constant