RISHAB LOKRAY
9357-3447

Fundamentals of Machine learning

Problem-1

i.i.d = $X = \{x_i\}_{i=1}^{N}$

$P(X/\mu)$ = Data likelihood

= Possions distribution

$$\left(P(x|\lambda) = \frac{\lambda^{x} e^{-\lambda}}{x!} \right)$$

I) Maximum Likelihood estimator: (MLE).

By MLE method we need to max. the data likelihood @ λ

$$\therefore \arg\max_{\lambda} P(X|\lambda) \quad ?$$

We know $P(X|\lambda)$ can be written as

$$P(X|\lambda) = P(x_1|\lambda) P(x_2|\lambda) \cdots$$

$$= \prod_{n=1}^{N} P(x_n|\lambda)$$

$\because$ as i.i.d

$$= \prod_{n=1}^{N} \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$$

→ Applying trick ~~on~~ (log on both sides)

$$\sum_{n=1}^{N} \ln \left(\frac{\lambda^{x_n} e^{-\lambda}}{x_n!} \right)$$

to find max value @ λ we differentiate w.r.t to λ: $\frac{\partial P(x|\lambda)}{\partial \lambda}$

$$\sum_{n=1}^{N} \left[\log \left(\lambda^{x_n} e^{-\lambda} \right) - \log (x_n!) \right]$$

$$\sum_{n=1}^{N} \left[-\lambda + x_n \log \lambda - \log (x_n!) \right]$$

$$\frac{\partial P(x|\lambda)}{\partial \lambda} \Rightarrow -N + \sum_{n=1}^{N} \frac{x_n}{\lambda} - 0$$

$$\therefore N = \frac{\sum_{n=1}^{N} x_n}{\lambda}$$

$$\boxed{\lambda = \frac{\sum_{n=1}^{N} x_n}{N}}$$

(II) Maximum A - posteriori : (MAP)

$$\underset{\lambda}{\arg\max}\; P(X/\lambda)\, P(\lambda)$$

i.e maximize the Posteriori.

~~to~~

$$\text{Prior} = P(\lambda/_{\alpha\beta}) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta}$$

$$\text{data likelihood} = P(X/\lambda) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

$$\Gamma(\alpha) = (\alpha-1)!$$

As it is i.i.d.

$$P(X/\lambda) = \prod_{n=1}^{N} \frac{\lambda^{x_n} e^{-\lambda}}{x_n!} = \text{data likelihood}$$

$$\text{Posteriori} = \left(\prod_{n=1}^{N} \frac{\lambda^{x_n} e^{-\lambda}}{x_n!}\right) \underbrace{\frac{\beta^{\alpha}}{(\alpha-1)!}}_{\text{constant}} \lambda^{\alpha-1} e^{-\lambda\beta}$$

$$\propto \left(\prod_{n=1}^{N} \left[\frac{\lambda^{x_n} e^{-\lambda}}{x_n!}\right] \lambda^{\alpha-1} e^{-\lambda\beta}\right)$$

$$\sum_{n=1}^{N}(-\lambda) + \sum_{n=1}^{N} x_n \log(\lambda) - \log(x_n!) + (\alpha-1)\log(\lambda) - \lambda\beta$$

$$\sum_{n=1}^{N}(-\lambda) - \log(x_n!) = \lambda\beta + \log(\lambda)\left((1-\alpha) - \sum_{n=1}^{N} x_n\right)$$

$$\frac{d\,\text{Posterior}}{d\lambda} \Rightarrow 0 \quad -N - 0 = \beta + \frac{1}{\lambda}\left(1-\alpha - \sum_{n=1}^{N} x_n\right)$$

$$\therefore \cancel{-N}$$

$$(-N-\beta) = \frac{1}{\lambda}\left(1-\alpha - \sum_{n=1}^{N} x_n\right)$$

$$\boxed{\lambda_{MAP} = \frac{\left(\sum_{n=1}^{N} x_n\right) + \alpha - 1}{N+\beta}}$$

Problem-3

I → The samples of X are i.i.d. ie.

$$X = \{x_i\}_{i=1}^{N} \quad \& \quad \{x_j\}_{j=N-10}^{N} = \text{Missing}$$

if $N = 50$

$$\{x_1 \; x_2 \; x_3 \; \ldots\ldots \; x_{39} \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; 0\}$$
$$x_{40} \; x_{41} \; \ldots\ldots \; x_{50}.$$

As the Value from X_{N-10} to X_N are missing we can ~~use~~ assume the values are censored to 0

for observed data likelihood: $P(x_i|\theta)$

$$L^0 = \prod_{i=1}^{N-10} P(x_i|\theta) \times \prod_{j=N-10}^{N} \int_a^{\infty} P(x_j|\theta)$$

- As values are missing we cannot just substitute value of 0 ~~as~~ as we will skew the results.
- we need to guess the ~~data~~ probabilities of each of the data points whose magnitudes we also not know.

I We can use a hidden variable z where we assume we know the values of the missing data points $z_j \ldots N$.

$$L^c = \prod_{i=1}^{N-10} P(x_i/\theta) \prod_{j=N-10}^{N} P(z_j/\theta).$$

if we have the z's we can find the Max of L^c and find θ values.

Problem 2