CS4641 HW2

Jordan Shartar
jsharter6
collaborators: Graeme Sharpe, Nicholas Loprinzo, Aishwarya Chitnis

1) Maximum Likelihood Estimation

a) Poisson distribution

$$P(x=k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (k=0,1,2,\dots)$$

Likelihood Function:

$$= \prod_{n=1}^{N} f_x(X_n;\lambda) = \prod_{n=1}^{N} \frac{e^{(-\lambda)}\lambda^{K_n}}{K_n!}$$

Log Likelihood:

$$= \log\left(\prod_{n=1}^{N} \frac{e^{(-\lambda)}\lambda^{K_n}}{K_n!}\right) = \sum_{n=1}^{N} \log\left(\frac{e^{-\lambda}\lambda^{K_n}}{K_n!}\right)$$

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

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

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

MLE λ:

$$\lambda_{MLE} = \underset{\lambda}{\arg\max}\; -N\lambda - \sum_{n=1}^{N} \log(K_n!) + \log(\lambda)\sum_{n=1}^{N} K_n$$

$$\frac{d}{d\lambda}\left(-N\lambda - \sum_{n=1}^{N} \log(K_n!) + \log(\lambda)\sum_{n=1}^{N} K_n\right)$$

$$= -N + \frac{1}{\lambda}\sum_{n=1}^{N} K_n = 0$$

$$\boxed{\lambda_{MLE} = \frac{1}{N}\sum_{n=1}^{N} K_n}$$

b) $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

Likelihood:

$$= \prod_{n=1}^{N} f_x(x_n ; \lambda) = \prod_{n=1}^{N} \lambda e^{(-\lambda x_n)}$$

$$= \lambda^N e^{(-\lambda \sum_{n=1}^{N} x_n)}$$

Log-Likelihood:

$$= \log(\lambda^N e^{(-\lambda \sum_{n=1}^{N} x_n)})$$

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

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

MLE λ:

$$\lambda_{MLE} = \underset{\lambda}{\operatorname{argmax}}\ N \log(\lambda) - \lambda \sum_{n=1}^{N} x_n$$

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

$$= \frac{N}{\lambda} - \sum_{n=1}^{N} x_n = 0$$

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

c) $N(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Likelihood:

$$= \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \qquad \star\ \frac{1}{\sigma\sqrt{2\pi}} = (2\pi\sigma^2)^{-1/2}$$

$$= (2\pi\sigma^2)^{-N/2} \exp\left(\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2\right)$$

Log-Likelihood:

$$\log\left((2\pi\sigma^2)^{-N/2} \exp\left(\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2\right)\right)$$

$$= \log\left((2\pi\sigma^2)^{-N/2}\right) + \log\left(\exp\left(\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2\right)\right)$$

$$= \frac{-N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2$$

$$= \frac{-N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2$$

MLE μ & σ^2:

$$\mu_{MLE} = \underset{\mu}{\operatorname{argmax}} -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2$$

$$\frac{d}{d\mu}\left(-\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n-\mu)^2\right)$$

$$= \frac{1}{\sigma^2} \sum_{n=1}^{N} (x_n-\mu) \quad = \frac{1}{\sigma^2}\left(\sum_{n=1}^{N}(x_n) - N\mu\right)$$

$$= \sum_{n=1}^{N} x_n - N\mu = 0$$

$$\mu_{MLE} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\sigma^2_{MLE} = \frac{d}{d\sigma^2}\left(-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right)$$

$$= \frac{-N}{2\sigma^2} - \left[\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right]\frac{d}{d\sigma^2}\left(\frac{1}{\sigma^2}\right)$$

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

$$= \frac{-N}{2\sigma^2} + \left[\frac{1}{2(\sigma^2)^2}\sum_{n=1}^{N}(x_n-\mu)^2\right]$$

$$= \frac{1}{2\sigma^2}\left[\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - N\right] = 0$$

$$\sum_{n=1}^{N}(x_n-\mu)^2 = N\sigma^2$$

$$\boxed{\sigma^2_{MLE} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2}$$

2] Chance of Karl Die

- Let $x^{(n)}$ be n^{th} sequence of M dice rolls

$$x^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots, x_M^{(n)})$$

- takes up to 12 values: $x_i^{(n)} \in \{1, \ldots, 12\}$
- Independent from other sequences
- dependent for current sequence

$$y^{(n)} = \begin{cases} 1 & \text{if Karl dice is used} \\ 0 & \text{if regular dice is used} \end{cases}$$

$$D = (x, y) = \{(x^{(n)}, y^{(n)})\}_{n=1}^{N}$$

a) Show Likelihood $P(D|h) = \prod_{n=1}^{N} [\pi_1 K(x^{(n)})]^{y(n)} [\pi_0 R(x^{(n)})]^{1-y(n)}$

where: $\pi_i = P(y=i)$

$K(x^{(n)}) = p(x^{(n)} | y^{(n)}=1)$

$R(x^{(n)}) = p(x^{(n)} | y^{(n)}=0)$

$(y^{(n)}=1)$ $P(D|h) = P(x^{(n)}, y^{(n)} | h) = \underbrace{P(x^{(n)} | y^{(n)}, h)}_{P(x^{(n)} | y^{(n)}=1)} \underbrace{P(y^{(n)})}_{\pi_1}$

$$= [\pi_1 K(x^{(n)})]^{y(n)}$$

$(y^{(n)}=0)$, $P(D|h) = P(x^{(n)}, y^{(n)} | h) = \underbrace{P(x^{(n)} | y^{(n)}=0, h)}_{P(x^{(n)} | y^{(n)}=0)} \underbrace{P(y^{(n)}=0)}_{\pi_0}$

$$= [\pi_0 R(x^{(n)})]^{1-y(n)}$$

$$\Rightarrow P(D|h) = \prod_{n=1}^{N} [\pi_1 K(x^{(n)})]^{y(n)} \cdot [\pi_0 R(x^{(n)})]^{1-y^{(n)}}$$

- The two steps above were the normalization of $P(D|h)$ separated to see each part's (Karl's Dice vs. Regular)

b) Show log-Likelihood is:

$$\log P(D|h) = \sum_{n=1}^{N} y^{(n)} \left[\log \pi_1 + \log K(x_1^{(n)}) + \sum_{i=2}^{M} \log K(x_i^{(n)} | x_{i-1}^{(n)})\right]$$

$$+ \sum_{n=1}^{N} (1-y^{(n)}) \left[\log \pi_0 + \sum_{i=1}^{M} \log R(x_i^{(n)})\right]$$

where: $K(x_1^{(n)}) = P(X_1^{(n)} | y^{(n)}=1)$

$K(x_i^{(n)} | x_{i-1}^{(n)}) = P(X_i^{(n)} | X_{i-1}^{(n)}, y=1)$

$R(x_i^{(n)}) = P(X_i^{(n)} | y^{(n)}=0)$

$$\log P(D|h) = \log\left(\prod_{n=1}^{N} [\pi_1 K(x^{(n)})]^{y^{(n)}} \cdot [\pi_0 R(x^{(n)})]^{1-y^{(n)}}\right)$$

$$= \sum_{n=1}^{N} y^{(n)} [\log \pi_1 + \log K(x^{(n)})] + \sum_{n=1}^{N} (1-y^{(n)})$$

$$\cdot [\log \pi_0 + \log R(x^{(n)})]$$

- Because Karls Dice is dependent on the last roll after the first is completed, Then you separate the $\sum_{n=1}^{N} y^{(n)} [\log \pi_1 + \log K(x^{(n)})]$ specifically the $\log K(x^{(n)})$ into the following:

$$\Rightarrow \log K(x^{(n)}) = \underbrace{\log P(x_1^{(n)} | y^{(n)} = 1)}_{K(x_1^{(n)})} + \sum_{i=2}^{M} \log P(x_i^{(n)} | x_{i-1}^{(n)}, y=$$

$$\Rightarrow \log P(D|h) = \sum_{n=1}^{N} y^{(n)} [\log \pi_1 + \log(K(x_1^{(n)})) +$$

$$\sum_{i=2}^{M} \log K(x_i^{(n)} | x_{i-1}^{(n)})] + \sum_{n=1}^{N} (1-y^{(n)})$$

$$\cdot [\log \pi_0 + \sum_{i=1}^{M} \log R(x_i^{(n)})]$$

c) optimization constraints for:

π_i, for $i = 1, 2$

$K(v)$ for $v = 1, \ldots, 12$

$K(v|v')$ for $v = 1, \ldots, 12$, $v' = 1, \ldots, 12$

$R(v)$ for $v = 1, \ldots, 12$

π_i : $\pi_1 + \pi_2 = 1$

$\pi_i - \pi_2 - 1 = 0$

$$\boxed{\sum_{i=1}^{K} \pi_i - 1 = 0}$$

$K(v)$: $K(x_i^{(n)}) = P(x_i^{(n)} \mid y^{(n)} = 1)$

$$\boxed{\sum_{v=1}^{V} K(v) - 1 = 0}$$

$K(v|v')$: $K(x_i^{(n)} \mid x_{i-1}^{(n)}) = P(x_i^{(n)} \mid x_{i-1}^{(n)}, Y=1)$

$$\boxed{\sum_{i=2}^{M} \sum_{v=1}^{V} \sum_{v'=1}^{V'} K(v_i \mid v'_{i-1}) - 1 = 0}$$

$R(v)$: $R(x^{(n)}) = P(x^{(n)} \mid y^{(n)} = 0)$

$$\boxed{\sum_{v=1}^{V} R(v) - 1 = 0}$$

d) Lagrangian form :

π_i : Loglikelihood() + $-\lambda\left(\sum_{i=1}^{K} \pi_i - 1\right)$

$K(v)$: loglikelihood() $- \lambda_1 \left(\sum_{v} K(v_1) - 1\right)$

$K(v|v')$: loglikelihood() $- \sum_{i=2}^{M} \lambda_i \left(\sum_{v} \sum_{v'} K(v|v') - 1\right)$

$R(v)$: Loglikelihood() $- \lambda\left(\sum_{v} (R(v)) - 1\right)$

e) Show expanded log-likelihood

$$\log P(D|h) = \sum_{n=1}^{N} Y^{(n)} \Big[\log \pi_1 + \sum_{v} I(X_1^{(n)}=v) \log K(X_1^{(n)}=v)$$
$$+ \sum_{i=2}^{M} \sum_{v} \sum_{v'} I(X_i^{(n)}=v, X_{i-1}^{(n)}=v') \log K(X_i^{(n)}=v \mid X_{i-1}^{(n)}=v') \Big]$$
$$+ \sum_{n=1}^{N} (1-Y^{(n)}) \Big[\log \pi_0 + \sum_{i=1}^{M} \sum_{v} I(X_i^{(n)}=v) \log R(X_i^{(n)}=v) \Big]$$

where: $I(a=b) = \begin{cases} 1 & \text{if } a=b \\ 0 & \text{otherwise} \end{cases}$

$$I(a=b, c=d) = \begin{cases} 1 & \text{if } a=b \text{ and } c=d \\ 0 & \text{otherwise} \end{cases}$$

$\sum_{v}$ is $\sum_{v \in \{1, \dots, 12\}}$

$\sum_{v'}$ is $\sum_{v' \in \{1, \dots, 12\}}$

Note: $\prod_{n=1}^{N} K(Y^{(n)}) = \prod_{n=1}^{N} \prod_{v=1}^{V} K(Y^{(n)}=v)^{I(Y^{(n)}=v)}$

$$\log \prod_{n=1}^{N} K(Y^{(n)}) = \sum_{n=1}^{N} \sum_{v=1}^{V} I(X^{(n)}=v) \log K(Y^{(n)}=v)$$

$$\Rightarrow \log P(D|h) = \sum_{n=1}^{N} Y^{(n)} \Big[\log \pi_1 + \log K(v) + \sum_{i=2}^{M} \log K(v|v') \Big]$$
$$+ \sum_{n=1}^{N} (1-Y^{(n)}) \Big[\log \pi_0 + \sum_{i=1}^{M} \log R(v) \Big]$$

$$\boxed{\begin{aligned} &= \sum_{n=1}^{N} Y^{(n)} \Big[\log \pi_1 + \sum_{v} I(X_1^{(n)}=v) \log K(X_i^{(n)}=v) + \\ &\quad \sum_{i=2}^{M} \sum_{v} \sum_{v'} I(X_i^{(n)}=v, X_{i-1}^{(n)}=v') \log K(X_i^{(n)}=v \mid X_{i-1}^{(n)}=v') \Big] \\ &+ \sum_{n=1}^{N} (1-Y^{(n)}) \Big[\log \pi_0 + \sum_{i=1}^{M} \sum_{v} I(X_i^{(n)}=v) \log R(X_i^{(n)}=v) \Big] \end{aligned}}$$

6) Show MLE π_1, $\pi_1 = \frac{1}{N}\sum_{n=1}^{N} y^{(n)}$

$$\frac{d}{d\pi_1}\log(D|h) = \sum_{n=1}^{N} y^{(n)}\left[\frac{1}{\pi_1}\right] + \sum_{n=1}^{N}(1-y^{(n)})\left[\frac{1}{1-\pi_1}\right] = 0$$

$$\sum_{n=1}^{N} y^{(n)}(1-\pi_1) = \sum_{n=1}^{N}(1-y^{(n)})\pi_1$$

$$\sum_{n=1}^{N} y^{(n)} - \cancel{\sum_{n=1}^{N} y^{(n)}\pi_1} = \underbrace{\sum_{n=1}^{N}\pi_1}_{N\pi_1} - \cancel{\sum_{n=1}^{N} y^{(n)}\pi_1}$$

$$\boxed{\pi_1 = \frac{1}{N}\sum_{n=1}^{N} y^{(n)}}$$

6) Show MLE $k(v)$, $k(v) = \frac{1}{N_1}\,\mathrm{count}_1(v, y=1)$

where: • N_1 is number of datapoints where $y=1$

• $\mathrm{count}_1(v, y=1)$ is number of datapoints with label $y=1$ where feature 1 have value v

$$\cancel{\frac{d}{dk(v)}\log P(D|h) = \sum_{n=1}^{N} y^{(n)}\left[\frac{\sum_v I(x_1^{(n)}=v)}{k(v)}\right] - \lambda_1 = 0}$$

$$\cancel{\sum_{n=1}^{N} y^{(n)} \quad \lambda_1 k(v) = \mathrm{count}(v, y=1)}$$

$$\cancel{\sum_{y^{(n)}=1} \quad = N_1}$$

$$\cancel{\boxed{k(v) = \frac{1}{N_1}\mathrm{count}(v, y=1)}}$$

$$\frac{d}{dk(v)}\log P(D|h) = \sum_{n=1}^{N} y^{(n)}\frac{1}{k(v)}\,\mathrm{count}_1(v, y=1) = 0$$

$$\frac{k(v)}{\mathrm{count}_1(v, y=1)} = \frac{1}{\sum_N y^{(n)}} \Rightarrow \frac{k(v)}{\mathrm{count}_1(v, y=1)} = \frac{1}{N_1}$$

$$\boxed{k(v) = \frac{1}{N_1}\,\mathrm{count}_1(v, y=1)}$$

h) MLE $K(v|v')$, $count_i(v, v', y^{(n)}=1)$ is useful

$$\cancel{\frac{d}{d(K(v|v'))} \log P(D|h) = \sum_{n=1}^{N} (y^{(n)}+1)[a + 0 + count_i(v, v', y^{(n)}=1)]}$$

$$\cancel{\sum_{i=2}^{M} \frac{1}{K(v|v')}}$$

$$\frac{d}{d(K(v|v'))} \log P(D|h) = \sum_{n=1}^{N} y^{(n)} \sum_{i=2}^{M} count_i(v, v', y^{(n)}=1) \cdot \frac{1}{K(v|v')} - \sum_{i=2}^{M} \lambda_i \sum_{v} \sum_{v'} 1 = 0$$

$$N_1 \sum_{i=2}^{M} count_i(v, v', y^{(n)}=1) = K(v|v') \sum_{i=2}^{M} \lambda_i \sum_{v} \sum_{v'} 1$$

Note: $\sum_{v} \sum_{v'} 1 = V V'$

$$\boxed{K(v|v') = N_1 \sum_{i=2}^{M} count_i(v, v', y^{(n)}=1) \cdot \frac{1}{\sum_{i=2}^{M} \lambda_i V V'}}$$

I) MLE $R(v)$

$$\frac{d}{dR(v)} \log p(D|h) = \sum_{n=1}^{N} (1 - y^{(n)}) \sum_{i=1}^{M} count_i(v, y=0) - \lambda \sum_{v} 1 = 0$$

$$\underbrace{\left(\sum_{n=1}^{N} 1 - \sum_{n=1}^{N} y^{(n)}\right)}_{N_0} \left(\sum_{i=1}^{M} count_i(v, y=0) \cdot \frac{1}{R(v)}\right) = \lambda V$$

$$\Rightarrow N_0 \sum_{i=1}^{M} count_i(v, y=0) = \lambda V R(v)$$

$$\boxed{R(v) = \frac{N_0 \sum_{i=1}^{M} count_i(v, y=0)}{\lambda V}}$$