Group Members:

- Nabil Arrouss (qup85ses)
- Ahmad Pamir Sahak (las06weq)
- Kohei Fujimura (koj66veq)
- Heydar Bagloo (yeq42hip)

Exercise 1:

$P(H | E_1 = e_1, E_2 = e_2)$ for $H \in \{1, \dots k\}$

$$P(H | e_1, e_2) = \frac{P(H, e_1, e_2)}{P(e_1, e_2)} = \frac{P(H) P(e_1, e_2 | H)}{P(e_1, e_2)}$$

ii) $P(e_1, e_2), P(H), P(e_1, e_2 | H)$, $P(H, e_1, e_2) = P(H) P(e_1, e_2 | H)$

$$\Rightarrow P(H | e_1, e_2) = \frac{P(H) P(e_1, e_2 | H)}{P(e_1, e_2)}$$

$$\Rightarrow P(e_1, e_2) = \sum_{h=1}^{k} P(H) P(e_1, e_2 | H)$$

iv) $P(H | e_1, e_2) = \frac{P(H, e_1, e_2)}{P(e_1, e_2)}$

$$\Rightarrow P(e_1, e_2) = \sum_{H=1}^{k} P(H, e_1, e_2)$$

$$\Rightarrow P(H | e_1, e_2) = \frac{P(H, e_1, e_2)}{\sum_{H=1}^{k} P(H, e_1, e_2)}$$

ii and iv are sufficient.

b) $E_1 \perp E_2 | H$ (conditional independence)

$$P(e_1, e_2 | H) = P(e_1 | H) P(e_2 | H)$$

$\Rightarrow P(H | e_1, e_2) = \frac{P(H) P(e_1 | H) \cdot P(e_2 | H)}{P(e_1, e_2)}$, $P(e_1, e_2) = \sum_H P(H) P(e_1 | H) P(e_2 | H)$

i) We can form the numerator $P(H) \cdot P(e_1 | H) \cdot P(e_2 | H)$ and the denominator is given so we can compute the posterior.

– sufficient under the independence assumption.

ii) sufficient without independence.

iii) Using conditional independence we can compute $P(e_1, e_2) = \sum_H P(H) P(e_1 | H) P(e_2 | H)$ so we can compute numerator and denominator and posterior.

– sufficient under the independence assumption.

CamScanner

iv; Sufficien (we know from Part a)

Exercise 2:

a; $X_1, X_2, X_3 \in \{0,1\}$, X_1, X_2 are independent fair coin

$P(X_1=0) = P(X_1=1) = \frac{1}{2}$, $P(X_2=0) = P(X_2=1) = \frac{1}{2}$, $X_1 \perp X_2$

$X_3 = X_1 \oplus X_2 \Rightarrow X_3 = 0$; X_1 and X_2, $X_3 = 1$, if $X_1 \neq X_2$

each X_i $P(X_i=0) = P(X_i=1) = \frac{1}{2}$

X_1 and X_3 for example:

$$P(X_1=0, X_3=0) = P(0,0,0) = \frac{1}{4}.$$

$$\Rightarrow \quad P(X_1=0)\,P(X_3=0) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

But

$$P(X_1=0, X_2=0, X_3=0) = \frac{1}{4}$$

$$P(X_1=0)\,P(X_2=0)\,P(X_3=0) = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}$$

Since $\frac{1}{4} \neq \frac{1}{8}$ so $P(X_1, X_2, X_3) \neq P(X_1)\cdot P(X_2)\cdot P(X_3)$

b;

assume that $X_1 \ldots X_n$ are mutually independent by:

$$P(X_1 = x_1 \ldots X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$\Rightarrow P(X_i = x_i, X_j = x_j) = \sum_{x_k,\, k \neq i,j} P(X_1 = x_1 \ldots X_n = x_n) = \sum_{x_k, k\neq i,j} \prod_{m=1} P(X_m, x_m)$$

since

$$P(X_i = x_i, X_j = x_j) = \left(P(X_i = x_i)\,P(X_j = x_j)\right) \prod_{k \neq i,j} \left(\sum_{x_k} P(X_k = x_k)\right)$$

and $\forall k$, $\sum_{x_k} P(X_k = x_k) = 1$

So $P(X_i = x_i, X_j = x_j) = P(X_i = x_i)\,P(X_j = x_j)$.

c) $P(X_1 | X_2) = P(X_1)$; suppose X_1, X_2 independent

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)\,P(X_2 = x_2)$$

for all x_1, x_2, if $P(X_2 = x_2) > 0$ then:

$$P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} = \frac{P(X_1 = x_1)\cdot \cancel{P(X_2 = x_2)}}{\cancel{P(X_2 = x_2)}}$$

$\Rightarrow P(X_1 = x_1)$

Exercise 3:

$$\text{Bias}[\hat{\theta}] = E[\hat{\theta}] - \theta^*$$

$$\text{Var}[\hat{\theta}] = E[(\hat{\theta} - E(\hat{\theta}))^2]$$

$$MSE = E[(\hat{\theta} - \theta^*)^2].$$

$$\Rightarrow E[(\hat{\theta} - \theta^*)^2] \quad , \quad \hat{\theta} - \theta^* = (\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta^*)$$

$$\Rightarrow (\hat{\theta} - \theta)^2 = [(\hat{\theta} - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta^*]^2$$

$$\Rightarrow (\hat{\theta} - E[\hat{\theta}])^2 + 2(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta^*) + (E[\hat{\theta}] - \theta)^2$$

$$\Rightarrow E[(\hat{\theta} - \theta^*)^2] = E[(\hat{\theta} - E[\hat{\theta}])]^2 + 2E[(\hat{\theta} - E[\hat{\theta}])(E[\hat{\theta}] - \theta^*)] + E[(E[\hat{\theta}] - \theta)^2]$$

$$E[(\hat{\theta} - E[\hat{\theta}])^2] = \text{Var}[\hat{\theta}]$$

$$\cancel{2(E[\hat{\theta}] - \theta^*) E[\hat{\theta} - E[\hat{\theta}]]}. \quad , \quad E[\hat{\theta} - E[\hat{\theta}]] = 0$$

$$E[E[\hat{\theta}] - \theta^*)^2] = (E[\hat{\theta} - \theta^*)^2 = (\text{Bias}[\hat{\theta}])^2$$

$$\Rightarrow E[(\hat{\theta} - \theta^*)^2] = \text{Var}[\hat{\theta}] + (\text{Bias}[\hat{\theta}])^2$$

$$\Rightarrow MSE[\hat{\theta}] = \text{Var}[\hat{\theta}] + (\text{Bias}[\hat{\theta}])^2.$$

Ex 4:

a)
$$l = \ln L = \ln\left(\prod_{i=1}^{N} \mathcal{N}(x_i|\mu, \sigma^2)\right)$$

$$\ln\left(\prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2\left(\frac{x_i-\mu}{\sigma}\right)^2}\right)$$

$$= \sum_{i=1}^{N} \ln\left(\frac{1}{\sigma\sqrt{2\pi}} \, e^{-1/2\left(\frac{x_i,\mu}{\sigma}\right)^2}\right)$$

$$= \sum_{i=1}^{N} \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$= \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

b)
$$\frac{\partial l}{\partial \mu} = \frac{\partial l}{\partial \mu}\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\right)$$

$$= -\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N} -2(x_i-\mu)\right)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu) \stackrel{!}{=} 0$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i - \sum_{i=1}^{N}\mu$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i - N\mu = 0$$

The term is 0 if $\sum_{i=1}^{N} x_i - N\mu = 0$

$$\sum_{i=1}^{N} x_i - N\mu \stackrel{!}{=} 0$$

$$\sum_{i=1}^{N} x_i = N\mu \Rightarrow \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

d) $\text{bias}[\hat{\mu}] = E[\hat{\mu} - \mu^*]$

$$= E\left[\frac{1}{N}\sum_{i=1}^{N} x_i - \mu^*\right]$$

$$= E\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] - E[\mu^*]$$

$$= \frac{1}{N} E\left[\sum_{i=1}^{N} x_i\right] - E[\mu^*]$$

$$= \frac{1}{N}\sum_{i=1}^{N} E[x_i] - E[\mu^*]$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu^* - \mu^* = 0$$

e) $\text{Var}[\hat{\mu}] = \text{Var}\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right]$

$$= \frac{1}{N^2}\sum_{i=1}^{N} \text{Var}[x_i]$$

f) $\text{MSE}[\hat{\mu}] = \text{bias}[\hat{\mu}]^2 + \text{Var}[\hat{\mu}]$

$$= \left(\frac{1}{N}\sum_{i=1}^{N} E[x_i] - E[\mu^*]\right)^2 + \frac{1}{N^2}\sum_{i=1}^{N} \text{Var}[x_i]$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} E[x_i]^2 - 2\cdot\frac{1}{N}\sum_{i=1}^{N} E[x_i]\,E[\mu^*] + E[\mu^*]^2 + \frac{1}{N^2}\sum_{i=1}^{N} \text{Var}[x_i]$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N} E[x_i] + \sum_{i=1}^{N} \text{Var}[x_i]\right] - \frac{2}{N}\sum_{i=1}^{N} E[x_i]\,E[\mu^*]$$

Machine Learning - Exercise sheet 1

Exercise 5) $D = \{x_1, \dots, x_N\}$ & $P(x|\lambda) = \lambda e^{-\lambda x}$ $\quad x \geq 0$, $\lambda > 0$

Likelihood:

$$L(\lambda) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i} = \lambda^N \exp\left(-\lambda \sum_{i=1}^{N} x_i\right)$$

Log-likelihood:

$$\ell(\lambda) = \log L(\lambda) = N \log \lambda - \lambda \sum_{i=1}^{N} x_i$$

$$\frac{d\ell}{d\lambda} = \frac{N}{\lambda} - \sum_{i=1}^{N} x_i$$

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

$$\Rightarrow \hat{\lambda}_{MLE} = \frac{N}{\sum_{i=1}^{N} x_i} = \frac{1}{\bar{x}}$$

s.t. $\bar{x}$ is the sample mean.