CS182 HW02 2021533138

1. (a) According to the definition of kernel function

we have that $K(x_i, x_j) \equiv \langle \phi(x_i), \phi(x_j) \rangle$, then for $\forall x_i, x_j \in$ space $\mathbb{R}^d$

with property of dot product, we get that $K(x_i, x_j) \equiv \langle \phi(x_i), \phi(x_j) \rangle \equiv \langle \phi(x_j), \phi(x_i) \rangle$

$\equiv K(x_j, x_i)$, so we get that $K(x_i, x_j) \equiv K(x_j, x_i)$

(b) $K(x_i, x_j) = \cancel{\exp(-\frac{1}{2}\|x_i - x_j\|^2)}$ ~~to~~ $\exp(-\frac{1}{2}\|x_i - x_j\|^2)$

then we have $\|\phi(x_i) - \phi(x_j)\|^2 = \langle \phi(x_i), \phi(x_i) \rangle - 2\langle \phi(x_i), \phi(x_j) \rangle + \langle \phi(x_j), \phi(x_j) \rangle$

$= K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j)$

$= \exp(0) - 2\exp(-\frac{1}{2}\|x_i - x_j\|^2) + \exp(0)$

$= 2 - 2\exp(-\frac{1}{2}\|x_i - x_j\|^2)$

~~as $\|x_i - x_j\|^2 \geq 0$, then we have $\|\phi(x_i) - \phi(x_j)\|^2$~~

as we have that $\exp(-\frac{1}{2}\|x_i - x_j\|^2) > 0$

then $2 - 2\exp(-\frac{1}{2}\|x_i - x_j\|^2) < 2$ then we got

$2 - 2\exp(-\frac{1}{2}\|x_i - x_j\|^2) \leq 2$

(c) x_{far} is a test point that is far away from training instance x_i,

which is measured in the origin space, then we have that

$\|x_{far} - x_i\|^2 \to +\infty$, then ~~as $\langle \hat{w}, \phi(x) \rangle + \hat{w}_0 = \sum_{i \in SV} y_i \alpha_i K(x_i, x) + \hat{w}_0$~~

~~$= \sum_{i \in SV} y_i \alpha_i \exp(-\frac{1}{2}\|x_i - x_j\|^2) + \hat{w}_0$~~ $\exp(-\frac{1}{2}\|x_{far} - x_i\|^2) \to 0$

$\langle \hat{w}, \phi(x) \rangle + \hat{w}_0 = \sum_{i \in SV} y_i \alpha_i K(x_{far}, x_i) + \hat{w}_0 = \sum_{i \in SV} y_i \alpha_i \exp(-\frac{1}{2}\|x_{far} - x_i\|^2) + \hat{w}_0 = \hat{w}_0$

then as $\langle \hat{w}, \phi(x) \rangle + \hat{w}_0 = f(x_{far}; \alpha, \hat{w}_0)$, we have that

$f(x_{far}; \alpha; \hat{w}_0) \propto \hat{w}_0$.

2. (a) $p(x_1, x_2, \dots, x_n | \lambda) = \prod_{i=1}^{n} p(x_i | \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$, which is the like-hood function

then we get the log of the like-hood function

we get $\log p(x_1, \dots, x_n|\lambda) = \log(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}) = \log\frac{\lambda^{x_1} e^{-\lambda}}{x_1!} + \dots + \log\frac{\lambda^{x_n} e^{-\lambda}}{x_n!}$

$= \sum_{i=1}^{n} \log \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \sum_{i=1}^{n} (x_i \log(\lambda) - \lambda - \log(x_i!)) = -n\lambda + \sum_{i=1}^{n} (x_i \log(\lambda) - \log(x_i!))$

then we make derivation to λ. we have that

$$\frac{\partial \log p(x_1, \dots, x_n|\lambda)}{\partial \lambda} = \frac{\partial(-n\lambda + \sum_{i=1}^{n}(x_i \log(\lambda) - \log(x_i!)))}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

the let derivation equal to 0 , we have $\frac{1}{\lambda}\sum_{i=1}^{n} x_i = n$

then $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$, so $\hat{\lambda}$ is the maximum likelihood estimate of λ.

Then as $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$, then $E(\hat{\lambda}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}E(\sum_{i=1}^{n} x_i) = \frac{1}{n} n\lambda = \lambda$

so $E(\hat{\lambda}) = \lambda$, so is unbiased

(b) PDF:

$$p(\lambda | \alpha, \beta) = \frac{\beta^\alpha}{(\alpha-1)!} \lambda^{\alpha-1} e^{-\beta\lambda}$$

posterior distribution over λ.

$$P(\lambda | x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n|\lambda) P(\lambda)}{P(x_1, x_2, \dots, x_n)} = \frac{P(x_1, x_2, \dots, x_n|\lambda) P(\lambda)}{\int_0^{+\infty} P(x_1, \dots, x_n|\lambda) P(\lambda)\, d\lambda}$$

$$= \frac{(\prod_{i=1}^{n} P(x_i|\lambda)) P(\lambda)}{\int_0^{+\infty} (\prod_{i=1}^{n} P(x_i|\lambda)) P(\lambda) d\lambda} = \frac{(\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_0^{+\infty} (\prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda} = \frac{e^{-n\lambda} (\prod_{i=1}^{n} \lambda^{x_i}) \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_0^{+\infty} e^{-n\lambda} (\prod_{i=1}^{n} \lambda^{x_i}) \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda}$$

$$= \frac{e^{-n\lambda} e^{-\beta\lambda} \lambda^{\sum_{i=1}^{n} x_i} \lambda^{\alpha-1}}{\int_0^{+\infty} e^{-n\lambda} e^{-\beta\lambda} \lambda^{\sum_{i=1}^{n} x_i} \lambda^{\alpha-1} d\lambda} = \frac{e^{-(n+\beta)\lambda} \lambda^{((\sum_{i=1}^{n} x_i) + \alpha - 1)}}{\int_0^{+\infty} e^{-(n+\beta)\lambda} \lambda^{((\sum_{i=1}^{n} x_i) + \alpha - 1)} d\lambda}$$

$$= \frac{e^{-(n+\beta)\lambda} \lambda^{((\sum_{i=1}^{n} x_i) + \alpha - 1)}}{\int_0^{+\infty} \frac{1}{\Gamma(\sum_{i=1}^{n} x_i + \alpha)} ((n+\beta)\lambda)^{\sum_{i=1}^{n} x_i + \alpha} e^{-(n+\beta)\lambda} \frac{d(n+\beta)\lambda}{(n+\beta)\lambda} \cdot \frac{\Gamma((\sum_{i=1}^{n} x_i) + \alpha)}{(n+\beta)^{\sum_{i=1}^{n} x_i + \alpha}}}$$ use the property of Gamma distribution

we have

$$= \frac{(n+\beta)^{(\sum_{i=1}^{n} x_i) + \alpha}}{\Gamma((\sum_{i=1}^{n} x_i) + \alpha)} e^{-(n+\beta)\lambda} \lambda^{((\sum_{i=1}^{n} x_i) + \alpha - 1)}$$

we have that

$P(\lambda | x_1, x_2, \dots, x_n) = \mathrm{Gamma}((\sum_{i=1}^{n} x_i) + \alpha, n+\beta)$

so $\lambda | x_1, \dots, x_n \sim \mathrm{Gamma}((\sum_{i=1}^{n} x_i) + \alpha, n+\beta)$

(c) Firstly, we have that

$$P(\lambda \mid x_1, \dots, x_n) = \cancel{(n+\beta)^{\sum_{i=1}^{n}}} \frac{(n+\beta)^{(\sum_{i=1}^{n} x_i)+\alpha}}{\Gamma((\sum_{i=1}^{n} x_i)+\alpha)} e^{-(n+\beta)\lambda} \lambda^{((\sum_{i=1}^{n} x_i)+\alpha-1)}$$

Firstly we get the log, that is

$$\ln(P(\lambda \mid x_1, \dots, x_n)) = \ln \frac{(n+\beta)^{(\sum_{i=1}^{n} x_i)+\alpha}}{\Gamma((\sum_{i=1}^{n} x_i)+\alpha)} + \ln e^{-(n+\beta)\lambda} + \ln \lambda^{((\sum_{i=1}^{n} x_i)+\alpha-1)}$$

then we make derivation

$$\frac{\partial \ln(P(\lambda \mid x_1, \dots, x_n))}{\partial \lambda} = 0 + (-(n+\beta)) + \frac{(\sum_{i=1}^{n} x_i)+\alpha-1}{\lambda} = 0$$

$$\text{so } \lambda = \frac{(\sum_{i=1}^{n} x_i)+\alpha-1}{n+\beta}$$

we get that ~~λ that~~ MAP of λ is $\frac{(\sum_{i=1}^{n} x_i)+\alpha-1}{n+\beta}$

3\. (a) Use D-separation

given D and F. as D and F are desendants of C

and we have that A, B, C is a head to head type.

we have that A.B are not independent.

so. A and B are not conditionally independent given D and F

(b) $P(D|CEG) \neq P(D|C)$

First we can consider that whether D and E are conditionally independent given C and that D and G conditionally independent given C.

Then ① D and E are conditionally independent given C as they are not connected

② D and G are not conditionally independent given C as they are connected

so we have D and E conditionally independent given C,

while D and G not given C.

so we have that $P(D|CEG) \neq P(D|C)$