HW1 　　　　　　　　　　　　　　 113064501 林峻霆

1. (10%) According to Eq. (2) and Eq. (3), please show that

$$\mathbb{E}[x_n x_m] = \mu^2 + I_{nm}\,\sigma^2. \tag{1}$$

x_n and x_m are two data points which sampled from a Gaussian distribution with mean μ, variance σ^2, and $I_{nm} = 1$ if $n = m$ otherwise $I_{nm} = 0$. Hence prove the results Eq. (4) and Eq. (5).

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x \mid \mu, \sigma^2)\, x\, dx = \mu. \tag{2}$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x \mid \mu, \sigma^2)\, x^2\, dx = \mu^2 + \sigma^2. \tag{3}$$

$$\mathbb{E}[\mu_{\mathrm{ML}}] = \mu. \tag{4}$$

$$\mathbb{E}[\sigma^2_{\mathrm{ML}}] = \left(\frac{N-1}{N}\right)\sigma^2. \tag{5}$$

① If $n = m$, we have

$$E[X_n X_m] = E[X_n^2] = \mu^2 + \sigma^2 \quad (\text{by Eq. }(3))$$

If $n \neq m$, we have $X_n \perp\!\!\!\perp X_m$, hence

$$E[X_n X_m] = E[X_n]\,E[X_m] = \mu^2$$

Therefore, we can conclude that $E[X_n X_m] = \mu^2 + I_{mn}\sigma^2$

② Given N samples $x_1, x_2 \cdots x_N$ the ML estimator of mean

$$\mu_{ML} = \underset{\mu}{\arg\max}\; N(x_1 \dots x_N \mid \mu, \sigma^2)$$

$$= \underset{\mu}{\arg\max} \prod_{i=1}^{N} N(x_i \mid \mu, \sigma^2)$$

$$= \underset{\mu}{\arg\max}\; -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2 - \frac{N}{2}\log 2\pi\sigma^2$$

$$= \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{N} (x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{N} (x_i - \mu)^2 = 0 \Rightarrow -\sum_{i=1}^{N} 2(x_i - \mu) = 0$$

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

Therefore we have

$$E[\mu_{ML}] = \frac{1}{N} \sum_{i=1}^{N} E[x_i] = \frac{1}{N} \cdot N\mu = \mu$$

③ Similarily, the ML estimator of variance is

$$\sigma^2_{ML} = \underset{\sigma^2_{ML}}{\operatorname{argmax}} \ N(x_1 \ldots x_N \mid \mu, \sigma^2)$$

$$= \underset{\sigma^2_{ML}}{\operatorname{argmax}} \ \underbrace{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2 - \frac{N}{2} \log 2\pi\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2} \left(\quad \right) = 0 \Rightarrow \frac{1}{2\sigma^4} \sum_{i=1}^{N} (x_i - \mu)^2 - \frac{N}{2\sigma^2} = 0$$

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

To meet the result of ②, the ML estimator of varaince

$$\sigma^2_{ML} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_{ML})^2$$

$$\Rightarrow E[\sigma^2_{ML}] = \frac{1}{N} \sum_{i=1}^{N} E\left[x_i^2 - 2x_i \frac{1}{N} \sum_{j=1}^{N} x_j + \frac{1}{N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} x_i x_j\right]$$

$$(\text{by } (2) \cdot (3)) = \frac{1}{N} \sum_{i=1}^{N} (\mu^2 + \sigma^2) - \frac{2}{N}(N\mu^2 + \sigma^2) + \frac{1}{N^2}(N^2\mu^2 + N\sigma^2)$$

$$= \frac{N-1}{N} \sigma^2$$

2. (10%) The uniform distribution for a continuous variable x is defined by

$$\mathrm{U}(x \mid a, b) = \frac{1}{b-a}, \quad a \leq x \leq b. \tag{6}$$

Verify that this distribution is normalized, and derive its mean and variance.

① $$\int_{-\infty}^{\infty} U(x \mid a, b)\, dx = \int_a^b \frac{1}{b-a}\, dx$$

$$= \frac{x}{b-a} \Big|_a^b$$

$$= 1$$

② $$E[X] = \int_a^b \frac{x}{b-a}\, dx$$

$$= \frac{1}{2(b-a)} x^2 \Big|_a^b$$

$$= \frac{b+a}{2}$$

③ $$E[X^2] = \int_a^b \frac{x^2}{b-a}\, dx$$

$$= \frac{1}{3(b-a)} x^3 \Big|_a^b$$

$$= \frac{1}{3}(b^2 + ab + a^2)$$

$$\Rightarrow Var(X) = E[X^2] - (E[X])^2$$

$$= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$$

$$= \frac{1}{12}(a^2+b^2) - \frac{1}{6}ab$$

$$= \frac{1}{12}(b-a)^2$$

3. (10%) The predictive distribution takes the form

$$p(t \mid \mathrm{x}, \mathbf{t}, \alpha, \beta) = \mathcal{N}\left(t \,\middle|\, \mathbf{m}_N^{\mathrm{T}}\phi(\mathrm{x}),\ \sigma_N^2(\mathrm{x})\right). \tag{7}$$

where the variance $\sigma_N^2(\mathrm{x})$ of the predictive distribution is given by

$$\sigma_N^2(\mathrm{x}) = \frac{1}{\beta} + \phi(\mathrm{x})^{\mathrm{T}}\mathbf{S}_N\,\phi(\mathrm{x}). \tag{8}$$

We know that as the size of the dataset increased, the uncertainty associated with the posterior distribution of the model parameters will be reduced. Make use of the matrix identity

$$\left(\mathrm{M} + \mathrm{v}\,\mathrm{v}^{\mathrm{T}}\right)^{-1} = \mathrm{M}^{-1} - \frac{(\mathrm{M}^{-1}\mathrm{v})\,(\mathrm{v}^{\mathrm{T}}\mathrm{M}^{-1})}{1 + \mathrm{v}^{\mathrm{T}}\mathrm{M}^{-1}\mathrm{v}}. \tag{9}$$

to show that the uncertainty $\sigma_N^2(\mathrm{x})$ associated with the linear regression function given by Eq. (8) satisfies

$$\sigma_{N+1}^2(\mathrm{x}) \le \sigma_N^2(\mathrm{x}). \tag{10}$$

The update rule for S_N is

$$S_N^{-1} = S_0^{-1} + \beta \sum_{i=1}^{N} \phi(x_i)\,\phi(x_i)^T$$

which results in the recurrent relation

$$S_{N+1}^{-1} = S_N^{-1} + \beta\,\phi(x_{N+1})\,\phi(x_{N+1})^T$$

$$\Rightarrow S_{N+1} = \left[S_N^{-1} + \beta\,\phi(x_{N+1})\,\phi(x_{N+1})^T\right]^{-1}$$

$$= S_N - \frac{\left[S_N \sqrt{\beta}\,\phi(x_{N+1})\right]\left[\sqrt{\beta}\,\phi(x_{N+1})^T S_N\right]}{1 + \beta\,\phi(x_{N+1})^T S_N\,\phi(x_{N+1})}$$

$$= S_N - \frac{S_N\,\phi(x_{N+1})\,\phi(x_{N+1})^T\,S_n}{1/\beta + \phi(x_{N+1})^T S_N\,\phi(x_{N+1})}$$

Therefore we have

$$\sigma_{N+1}^2(x) = 1/\beta + \phi(x)^T S_{N+1} \phi(x)$$

$$= 1/\beta + \phi(x)^T \left[S_N - \frac{S_N \phi(x_{N+1}) \phi(x_{N+1})^T S_n}{1/\beta + \phi(x_{N+1})^T S_N \phi(x_{N+1})} \right] \phi(x)$$

$$= \sigma_N^2(x) - \frac{\phi(x)^T S_N \phi(x_{N+1}) \phi(x_{N+1})^T S_N \phi(x)}{1/\beta + \phi(x_{N+1})^T S_N \phi(x_{N+1})}$$

$$= \sigma_N^2(x) - \frac{| \phi(x)^T S_N \phi(x_{N+1}) |^2}{1/\beta + \phi(x_{N+1})^T S_N \phi(x_{N+1})}$$

Since S_N is a positive semidefinite (PSD) matrix,

$$\frac{| \phi(x)^T S_N \phi(x_{N+1}) |^2}{1/\beta + \phi(x_{N+1})^T S_N \phi(x_{N+1})} \geq 0$$

Hence we can conclude that

$$\sigma_{N+1}^2(x) \leq \sigma_N^2(x)$$

4. (10%) The beta distribution, given by Eq. (11), is correctly normalized, so that Eq. (12) holds:

$$\mathrm{Beta}(\mu \mid a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, \mu^{a-1}\,(1-\mu)^{b-1}. \tag{11}$$

$$\int_0^1 \mu^{a-1}\,(1-\mu)^{b-1}\, d\mu = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}. \tag{12}$$

Make use of the result Eq. (12) to show that the mean and variance of the beta distribution Eq. (11) are given respectively by:

$$\mathbb{E}[\mu] = \frac{a}{a+b}. \tag{13}$$

$$\mathrm{var}[\mu] = \frac{a\,b}{(a+b)^2\,(a+b+1)}. \tag{14}$$

① $E[X] = \int_0^1 x \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} dx$

$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1} dx$

$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} = \frac{(a+b-1)!\ a!}{(a-1)!\ (a+b)!} = \frac{a}{a+b}$

② $E[x^2] = \int_0^1 x^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} dx$

$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)}$

$= \frac{(a+1)\,a}{(a+b+1)\,(a+b)}$

$\exists\ Var(X) = E[x^2] - (E[X])^2$

$= \frac{(a+1)\,a}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b)^2 (a+b+1)}$