Date: __/__/____

Machine learning - B555
Written Assignment 1

Prob 1.13

Given:

$$\sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2 \quad — (1.56)$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n \quad — (1.55)$$

To prove:

$$E(\sigma^2_{ML}) = \sigma^2 \text{ if } \mu_{ML} = \mu$$

Proof:

$$E(\sigma^2_{ML}) = E\left[\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2\right] \quad \left(\underset{\text{Given}}{\mu_{ML} = \mu}\right)$$

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

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

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

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

$$= \sigma^2 + \mu^2 - \frac{2}{N}\times\mu\times N\mu + \mu^2$$

#learnthesmarterway

Scanned by CamScanner

Date: __/__/____

$E(\sigma^2_{ML}) = \sigma^2 + 2\mu^2 - 2\mu^2$

$$\boxed{E(\sigma^2_{ML}) = \sigma^2}$$

Here, we use the following 2 properties of Expectation

$$E[x] = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)\, x\, dx = \mu$$

$$E[x^2] = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)\, x^2\, dx = \mu^2 + \sigma^2$$

Given:

Q.2 Poisson Distribution $\rightarrow P(x|\lambda) = \dfrac{e^{-\lambda} \times \lambda^x}{x!}$

Here, $E[x] = Var[x] = \lambda$

IID : $\{x_1, x_2, \ldots x_N\}$

Find:

(i) Max^m likelihood estimate of λ

(ii) Mean & Variance of maximal likelihood estimator

Solution:

$P(x_1|\lambda) = \dfrac{e^{-\lambda} \times \lambda^{x_1}}{(x_1)!} \longrightarrow P(x_n|\lambda) = \dfrac{e^{-\lambda} \times \lambda^{x_n}}{(x_n)!}$

$P(x_2|\lambda) = \dfrac{e^{-\lambda} \times \lambda^{x_2}}{(x_2)!}$

⋮

#learnthesmarterway

Scanned by CamScanner

Date: __/__/___

$$\text{Likelihood } (L) = \prod_{n=1}^{n} \frac{\exp(-\lambda) \cdot \lambda^{x_n}}{(x_n)!}$$

Taking Log

$$\ln(L) = \ln\left(\prod_{n=1}^{n} \frac{\exp(-\lambda) \cdot \lambda^{x_n}}{(x_n)!}\right)$$

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

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

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

$$\ln(L) = -n\lambda - \sum_{n=1}^{n} \ln(x_n!) + \ln\lambda \sum_{n=1}^{n} x_n \quad \text{--- (1)}$$

Differentiate (1) w.r.t λ & equate to 0

$$\frac{d \ln(L)}{d\lambda} = -n + \frac{1}{\lambda} \sum_{n=1}^{n} x_n = 0$$

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

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

#learnthesmarterway

Scanned by CamScanner

Date: __/__/___

(ii) The mean & variance of $\hat{\lambda}$

Mean: $E\hat{\lambda} = E\left[\frac{1}{n}\sum_{n=1}^{n} x_n\right]$

$= \frac{1}{n} \times E\left[\sum_{n=1}^{n} x_n\right]$

$= \frac{1}{\cancel{n}} \times \cancel{n} \times \lambda$ [Given that $E(x) = \lambda$]

$$\boxed{E\hat{\lambda} = \lambda}$$

Variance:

$Var(\hat{\lambda}) = Var\left(\frac{1}{n}\sum_{n=1}^{n} x_n\right)$

$= \frac{1}{n^2}\sum_{n=1}^{n} Var(x_n)$

$= \frac{1}{n^{\cancel{2}}} \times \cancel{n}\lambda$ [Given that $Var[x] = \lambda$]

$$\boxed{Var\hat{\lambda} = \frac{\lambda}{n}}$$

Problem 1.30 $p(x) = N(x|\mu, \sigma^2)$

$q(x) = N(x|m, s^2)$

Acc to Kullback-Leibler Divergence

$KL(p||q) = -\int p(x) \cdot \ln\left(\frac{q(x)}{p(x)}\right) dx$

#learnthesmarterway

Scanned by CamScanner

Date: __/__/____

$$= \int p(x) \cdot \ln\left(\frac{p(x)}{q(x)}\right) dx$$

$$p(x) = \sqrt{\frac{1}{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$q(x) = \frac{1}{\sqrt{2\pi s^2}} \cdot e^{-\frac{1}{2s^2}(x-m)^2}$$

$$\frac{p(x)}{q(x)} = \frac{s}{\sigma} \cdot \frac{\exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2\right\}}{\exp\left\{\frac{-1}{2s^2}(x-m)^2\right\}}$$

$$\cancel{\ln} \quad = \frac{s}{\sigma} \cdot \exp\left\{\frac{-1}{2\sigma^2}(x-\mu)^2 + \frac{1}{2s^2}(x-m)^2\right\}$$

$$\ln\left(\frac{p(x)}{q(x)}\right) = \ln\left(\frac{s}{\sigma}\right) + \left(\frac{1}{2s^2}(x-m)^2 - \frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$= \ln\left(\frac{s}{\sigma}\right) + \left[\frac{x^2+m^2-2xm}{2s^2} - \left\{\frac{x^2+\mu^2-2\mu x}{2\sigma^2}\right\}\right]$$

$$= \ln\left(\frac{s}{\sigma}\right) + x^2\left(\frac{1}{2s^2} - \frac{1}{2\sigma^2}\right) - \frac{2x}{\cancel{2}}\left(\frac{m}{\cancel{2}s^2} - \frac{\mu}{\sigma^2}\right) + \left(\frac{m^2}{2s^2} - \frac{\mu^2}{2\sigma^2}\right)$$

$$= \ln\left(\frac{s}{\sigma}\right) + \frac{\sigma^2 + (\mu-m)^2}{2s^2} - \frac{1}{2}$$

$$y = \frac{s}{\sigma}$$

#learnthesmarterway

Scanned by CamScanner

Date: __/__/____

~~KL (p||q) = ln x~~

$$KL(p||q) = \ln x + \frac{\sigma^2}{2s^2} - \frac{1}{2} + \frac{(\mu - m)^2}{2s^2}$$

$$KL(p||q) = \ln x + \frac{1}{2x^2} - \frac{1}{2} + \frac{(\mu - m)^2}{2s^2} \quad \text{--- } 1$$

$$\frac{d(KL)}{dx} = \frac{1}{x} - \frac{1}{x^3}$$

$$\frac{d(KL)}{dx} = \frac{1}{x}\left(1 - \frac{1}{x^2}\right)$$

Equating to 0

$$\frac{1}{x}\left(1 - \frac{1}{x^2}\right) = 0$$

$$1 = \frac{1}{x^2}$$

$$\boxed{x = \pm 1}$$ (Can't be negative)

$$KL(p||q) = \ln(1) + \frac{1}{2} - \frac{1}{2} + \frac{(\mu - m)^2}{2s^2}$$

$$\boxed{KL(p||q) = \frac{(\mu - m)^2}{2s^2}}$$

#learnthesmarterway

Scanned by CamScanner

5. Prob 2.38 Likelihood given by:

$$\boxed{p(x|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \cdot \exp\left\{\frac{-1}{2\sigma^2}\sum(x_n-\mu)^2\right\}}$$

Prior distribut^n given by:

$$\boxed{P(\mu) = \frac{1}{(2\pi\sigma_0^2)^{\frac{1}{2}}} \cdot \exp\left\{\frac{-1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}}$$

Posterior Distribut^n given by:

$$\boxed{P(\mu|\mathbf{x}) = \frac{1}{(2\pi\sigma_N^2)^{\frac{1}{2}}} \cdot \exp\left\{\frac{-1}{2\sigma_N^2}(\mu-\mu_N)^2\right\}}$$

Now,

Posterior $\propto$ Likelihood x Prior

$$\frac{1}{(2\pi\sigma^2)^{N/2}} \cdot \exp\left\{\frac{-1}{2\sigma^2}\sum(x_n-\mu)^2\right\} \times \frac{1}{(2\pi\sigma_0^2)^{\frac{1}{2}}} \cdot \exp\left\{\frac{-1}{2\sigma_0^2}(\mu-\mu_0)^2\right\}$$

$$\propto \frac{1}{(2\pi\sigma_N^2)^{\frac{1}{2}}} \cdot \exp\left\{\frac{-1}{2\sigma_N^2}(\mu-\mu_N)^2\right\}$$

Comparing exponent terms

$$\exp\left\{-\frac{(\mu-\mu_N)^2}{2\sigma_N^2}\right\} \propto \exp\left\{\frac{\sum(x-\mu)^2}{2\sigma^2}\right\} \times \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}$$

$$\Rightarrow \exp\left\{-\frac{(\mu-\mu_N)^2}{2\sigma_N^2}\right\} \propto \exp\left\{\frac{\sum(x-\mu)^2}{2\sigma^2} - \frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}$$

#learnthesmarterway

Scanned by CamScanner

Date: __/__/____

Comparing μ^2 terms on both sides

$$-\frac{1}{2\sigma_N^2} = -\frac{N}{2\sigma^2} - \frac{1}{2\sigma_0^2}$$

or $$\boxed{\frac{1}{\sigma_N^2} = \frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}} \rightarrow (2.142)$$

Now, comparing μ terms of exponent

$$\frac{\mu^2 + \mu_N^2 - 2\mu\cdot\mu_N}{2\sigma_N^2} \quad \alpha \sum_{n=1}^{n}\left(\frac{x^2 + \mu^2 - 2\mu x}{2\sigma^2}\right) + \frac{\mu^2 + \mu_0^2 - 2\mu\mu_0}{2\sigma_0^2}$$

$$\mu\left(\frac{\mu_N}{\sigma_N^2}\right) \quad \alpha \mu\left(-\frac{x}{\sigma^2} - \frac{\mu_0}{\sigma_0^2}\right)$$

$$\frac{\mu_N}{\sigma_N^2} = \frac{\sum x_n}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$$

$$\mu_N\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right) = \frac{N\cdot\mu_{ML}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$$

$$\mu_N\left(\frac{N\sigma_0^2 + \sigma^2}{\cancel{\sigma^2\sigma_0^2}}\right) = \frac{N\cdot\mu_{ML}\cdot\sigma_0^2 + \mu_0\cdot\sigma^2}{\cancel{\sigma^2\sigma_0^2}}$$

$$\boxed{\mu_N = \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\cdot\mu_{ML} + \frac{\sigma^2\cdot\mu_0}{N\sigma_0^2 + \sigma^2}} \rightarrow (2.141)$$

#learnthesmarterway

Scanned by CamScanner

Date: __/__/__

Problem 2.10 $\Gamma(x+1) = x\Gamma(x)$ — (1)

$$Dir(\mu|\alpha) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} \cdot \prod_{k=1}^{K} \mu_k^{\alpha_k - 1} \quad \text{— (2)}$$

Prove that:

(i) $$E[\mu_j] = \frac{\alpha_j}{\alpha_0}$$

$$E[\mu_j] = \int \mu_j \times Dir(\mu|\alpha)\, d\mu$$

$$E[\mu_j] = \int \mu_j \cdot \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)} \cdot \prod_{k=1}^{K} \mu_k^{\alpha_K - 1} d\mu$$

$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_K)} \cdot \int \mu_j \cdot \prod_{k=1}^{K} \mu_k^{\alpha_K - 1} d\mu$$

$$\cancel{\alpha_0 = \Sigma_k} \;\; " \cdot \int \prod_{\substack{k=1 \\ k \neq j}} \cdot \mu_k^{\alpha_k - 1} \cdot \mu_j^{\alpha_j + 1 - 1} d\mu_j$$

$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_k)} \cdot \prod_{\substack{k=1 \\ k \neq j}}^{M} \frac{\Gamma(\alpha_k)\cdots\Gamma(\alpha_j+1)}{\Gamma(\Sigma(\alpha_k)+1}$$

$$= \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_k)} \cdot \prod_{k=1}^{M} \frac{\Gamma(\alpha_k)\cdots\Gamma(\alpha_j+1}{\Gamma(\alpha_0)+1}$$

$$= \frac{\Gamma(\alpha_0) \cdot \Gamma(\alpha_j+1)}{\Gamma(\alpha_j) \cdot \Gamma(\alpha_0)+1} = \frac{\cancel{\Gamma(\alpha_0)} \cdot \cancel{\Gamma(\alpha_j)} \times \alpha_j}{\cancel{\Gamma(\alpha_j)} \times \alpha_0 \cdot \cancel{\Gamma(\alpha_0)}}$$

#learnthesmarterway

Scanned by CamScanner

Date: __/__/____

$$\boxed{E(\mu_j) = \frac{\alpha_j}{\alpha_0}}$$

#learnthesmarterway

Scanned by CamScanner

References For the Assignment

[1] https://www.youtube.com/watch?v=nfBNOWv1pgE
[2] https://github.com/GoldenCheese/PRML-Solution-Manual
[3] https://www.microsoft.com/en-us/research/wp-content/uploads/2016/05/prml-web-sol-2009-09-08.pdf
[4] http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf
[5] https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

Scanned by CamScanner