Yang, Ya-Chu
11571276
qwerthjkl45@gmail.com

Homework Assignment 2
Machine Learning 1, 17/18

2017-09-20

1 MAP solution for Linear Regression

1.

$$p(\boldsymbol{t}|\boldsymbol{w}) = \prod_{n=1}^{N} p(t_n|\phi_n, \boldsymbol{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w^T}\phi_n, \beta^{-1}) = \prod_{n=1}^{N} \frac{1}{(2\pi\beta^{-1})^{\frac{1}{2}}} exp[-\frac{1}{2\beta^{-1}}(t_n - \boldsymbol{w^T}\phi_n)^2]$$
$$= \frac{1}{(2\pi\beta^{-1})^{\frac{1}{2}}} exp[-\frac{1}{2\beta^{-1}}(\boldsymbol{t} - \Phi\boldsymbol{w})^T(\boldsymbol{t} - \Phi\boldsymbol{w})]$$

2.

$$p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}|0, \alpha^{-1}I) = \left(\frac{\alpha}{2\pi}\right)^{\frac{M}{2}} exp\{-\frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w}\}$$
$$ln(p(\boldsymbol{w})) = \frac{M}{2}ln(\frac{\alpha}{2\pi}) - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w}$$

3.

$$p(\boldsymbol{w}|\boldsymbol{t}, \alpha, \beta) = \frac{p(\boldsymbol{t}|\boldsymbol{w})p(\boldsymbol{w})}{p(\boldsymbol{t})} = \frac{[\prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w^T}\phi_n, \beta^{-1})]\mathcal{N}(\boldsymbol{w}|0, \alpha^{-1}I)}{\int\{\prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w^T}\phi_n, \beta^{-1})]\mathcal{N}(\boldsymbol{w}|0, \alpha^{-1}I)\}d\boldsymbol{w}}$$

4.

$$ln(p(\boldsymbol{w}|\boldsymbol{t}, \alpha, \beta)) = ln(p(\boldsymbol{t}|\boldsymbol{w})) + ln(p(\boldsymbol{w})) - ln(\int\{\prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w^T}\phi_n, \beta^{-1})]\mathcal{N}(\boldsymbol{w}|0, \alpha^{-1}I)\}d\boldsymbol{w})$$
$$=ln(\frac{1}{(2\pi\beta^{-1})^{\frac{N}{2}}}) - \frac{1}{2\beta^{-1}}\left\|\boldsymbol{t} - \Phi\boldsymbol{w}\right\|^2 + \frac{M}{2}ln(\frac{\alpha}{2\pi}) - \frac{\alpha}{2}\boldsymbol{w}^T\boldsymbol{w} - ln(\int\{\prod_{n=1}^{N} \mathcal{N}(t_n|\boldsymbol{w^T}\phi_n, \beta^{-1})]\mathcal{N}(\boldsymbol{w}|0, \alpha^{-1}I)\}d\boldsymbol{w})$$
$$=C - \frac{1}{2\beta^{-1}}\left\|\boldsymbol{t} - \Phi\boldsymbol{w}\right\|^2 - \frac{\alpha}{2}\left\|\boldsymbol{w}\right\|^2$$

5.

$$\partial\frac{ln(p(\boldsymbol{w}|\boldsymbol{t}, \alpha, \beta))}{\partial\boldsymbol{w}} = \frac{1}{2\beta^{-1}}(2\Phi)^T(\boldsymbol{t} - \Phi\boldsymbol{w}) - \frac{\alpha}{2}2\boldsymbol{w} = \frac{1}{\beta^{-1}}\Phi^T(\boldsymbol{t} - \Phi\boldsymbol{w}) - \alpha\boldsymbol{w} = 0$$
$$\Phi^T(\boldsymbol{t} - \Phi\boldsymbol{w}) - \beta^{-1}\alpha\boldsymbol{w} = 0$$
$$\Phi^T t = (\Phi^T\Phi + \beta^{-1}\alpha I)\boldsymbol{w}$$
$$\boldsymbol{w} = (\Phi^T\Phi + \beta^{-1}\alpha I)^{-1}\Phi^T t$$

2 Probability distributions, likelihoods, and estimators

2.1

Bernouilli: It is a kind of probability distribution takes value 1 or value 0 , which often being applied in yes/no question. For example, describing the outcome of flipping a coin, with $\theta = 0$ representing ”tailsänd $\theta = 1$ representing ”heads”.The probability distribution over θ can therefore be written in this form

Beta: It is a kind of probability distribution takes value between 0 to 1, which often being applied to model the behavior of random variables limited to intervals of finite length in a wide variety of disciplines.For example, describing a production plant produces items that have a probability X of being defective.

Poisson: It is a kind of probability distribution takes positive integer value, which often being applied to expresses the probability of a given number of events occurring in a fixed interval of time and/or space. For example, making forecasts about the number of customers or sales on certain days or seasons of the year by businessman.

Gamma: It is a kind of probability distribution takes positive real value, which is based on the interval which occurs between events when derived from it becomes the sum one or more than one exponentially distributed variables. For example, making forecasts about rainfalls.

Gaussian: It is a kind of probability distribution takes real value. For example, being applied in biology, like measuring size of living tissue.

Log-normal: It is a kind of probability distribution takes positive real value, and the random variable whose logarithm is normally distributed. This model can be applied on describing natural phenomena, like natural growth process.

2.2

1.

$$\text{single observation: } \rho^{r_t}(1-\rho)^{1-r_t}$$
$$\text{entire set of observations: } \prod_{t=1}^{N} \rho^{r_t}(1-\rho)^{1-r_t} = \rho^{n_1}(1-\rho)^{n_0}$$

2.

$$\boldsymbol{r} : \text{represent for the entire set of observations}$$
$$lnp(\boldsymbol{r}|\rho) = n_1 ln\rho + n_0 ln(1-\rho)$$

3.

$$\partial\frac{lnp(\boldsymbol{r}|\rho)}{\partial\boldsymbol{\rho}} = \frac{n_1}{\rho} - \frac{n_0}{1-\rho} = 0$$
$$\rho = \frac{n_1}{n_1+n_0} = \frac{207}{365} \simeq 0.57$$

4.

$$p(\rho|n_1,n_0,a,b) \propto p(\boldsymbol{r}|\rho)p(\rho|a,b)$$
$$p(\rho|n_1,n_0,a,b) \propto [\rho^{n_1}(1-\rho)^{n_0}][\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\rho^{a-1}(a-\rho)^{b-1}] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\rho^{n_1+a-1}(1-\rho)^{n_0+b-1}$$
$$\frac{\partial p(\rho|n_1,n_0,a,b)}{\partial\rho} = \frac{n_1+a-1}{\rho} - \frac{n_0+b-1}{1-\rho} = 0$$
$$\rho = \frac{n_1+a-1}{n_1+n_0+a+b-2}$$

5.

$$p(\rho|n_0,n_1,a,b) = \frac{p(\boldsymbol{r}|\rho)p(\rho|a,b)}{p(\boldsymbol{r})} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\rho^{n_1+a-1}(1-\rho)^{n_0+b-1}}{\int[p(\boldsymbol{r}|\rho)p(\rho|a,b)]d\rho}$$

6.

2.3

1.

$$\text{single observation: } \frac{\lambda^{d_t}e^{-\lambda}}{(d_t)!}$$
$$\text{entire set of observations: } \prod_{t=1}^{T}\frac{\lambda^{d_t}e^{-\lambda}}{(d_t)!} = e^{-T\lambda}\lambda^{n}\prod_{t=1}^{T}\frac{1}{(d_t)!}$$

2.

$$\text{log-likelihood} : ln\left[(e^{-T\lambda}\lambda^n\prod_{t=1}^{T}\frac{1}{(d_t)!})\right] = -T\lambda + nln\lambda - \sum_{t=1}^{T} ln((d_t)!)$$

3.

$$\partial\frac{-T\lambda + nln\lambda - \sum_{t=1}^{T} ln((d_t)!)}{\partial\lambda} = -T + \frac{n}{\lambda} = 0$$
$$\lambda = \frac{n}{T} = \frac{43}{14} \simeq 3.07$$

4.

$\boldsymbol{d}$: represent for the entire set of observations

$p(\lambda|n, T, a, b) \propto p(\boldsymbol{d}|\lambda)p(\lambda|a, b)$

$$\partial\frac{lnp(\lambda|n, T, a, b)}{\partial\lambda} = \partial\frac{[lnp(\boldsymbol{d}|\lambda) + lnp(\lambda|a, b)])}{\partial\lambda} = \partial\frac{-T\lambda + nln\lambda - \sum_{t=1}^{T} ln((d_t)!) + ln[\frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-b\lambda}]}{\partial\lambda}$$
$$= (-T - b) + \frac{n + a - 1}{\lambda} = 0$$
$$\lambda = \frac{n + a - 1}{T + b}$$

5.

$$p(\lambda|n, T, a, b) = \frac{p(\boldsymbol{d}|\lambda)p(\lambda|a, b)}{p(\boldsymbol{d})} = \frac{\prod_{t=1}^{T}[\frac{1}{(d_t)!}]\frac{b^a}{\Gamma(a)}e^{\lambda(-T-b)}\lambda^{n+a-1}}{\int[p(\boldsymbol{d}|\lambda)p(\lambda|a, b)]d\lambda}$$