Paul Ngouchet
BUID: U74701506

CS542: Machine Learning

Problem set 2.

I. Written Problems.

(a) Bishop 3.3.

If we define $R = \mathrm{diag}(r_1, \dots, r_N)$ to be a diagonal matrix containing the weighting coefficients

$$\Rightarrow E_D(w) = \frac{1}{2}(t - \Phi w)^T R (t - \Phi w)$$

$$dw = (\Phi^T R \Phi)^{-1} \Phi^T R t.$$

which becomes like (3.15) $R = I$

by comparing (3.104) with (3.10) - (3.12)

* r_n is a precision parameter (inverse variance) ~~Postr~~ regarding the point (x_n, t_n) which can either replace or scale β.

* or r_n could be ~~as~~ an effective number of replicated observations of (x_n, t_n). If we consider (3.104), ~~tt~~ It becomes clear if we consider (3.104) with r_n taking positive integer values, ~~at~~ even though it is also valid for any $r_n > 0$

3.11.

~~From~~ Based on (3.59) we have

$$\sigma^2_{N+1}(x) = \frac{1}{\beta} + \phi(x)^T S_{N+1} \phi(x)$$

where S_{N+1} is given by (1.31). with (1.31) and 3.110 we get.

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

$$= S_N - \frac{(S_N \phi_{N+1} \beta^{\frac{1}{2}})(\beta^{1/2} \phi_{N+1}^T S_N)}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}}$$

$$= S_N - \frac{\beta S_N \phi_{N+1} + \phi_{N+1}^T S_N}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}}$$

With 3.59, we can write (133) as.

$$\sigma^2_{N+1}(x) = \frac{1}{\beta} + \phi(x)^T \left(S_N - \frac{\beta S_N \phi_{N+1} \phi_{N+1}^T S_N}{1 + \beta \phi_{N+1}^T S_N \phi_{N+1}} \right) \phi(x)$$

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

Since S_N is positive define numerator and denominator will be non negative and positive.

hence $\sigma^2_{N+1}(x) \leq \sigma^2_N(x)$

(2)

3.14

$\alpha = 0 \Rightarrow S_N = (\beta \Phi^T \Phi)^{-1}$ (140)

$\psi(x) = V\phi(x)$ (141)

$V = M \times M$ Matrix

$\phi(x) = V^{-1}\psi(x)$

(141) and (3.16) $\Rightarrow \Psi = \Phi V^T$

$\Rightarrow \Phi = \Psi V^{-T}$ $\Rightarrow$ where $V^{-T} = (V^{-1})^T$

$\Psi^T \Psi = I$.

$(V^{-1})^T = (V^T)^{-1}$

From 140 $S_N = \beta^{-1}(\Phi^T \Phi)^{-1} = \beta^{-1}(V^{-T}\Psi^T \Psi V^{-1})^{-1}$

$= \beta^{-1} \epsilon V^T V$

The equivalent kernel:

$k(x, x') = \beta\phi(x)^T S_N \phi(x') = \phi(x)^T V^T V \phi(x')$

$= \psi(x)^T \psi(x')$

with $J = 1$

$\Rightarrow \sum_{n=1}^{N} \psi_i(x_n)\, \psi_1(x_n) = \sum_{n=1}^{N} \psi_i(x_n) = \delta_{i1}$

where $\psi(x) = 1$.

$$\Leftarrow) \sum_{n=1}^{N} k(x, x_n) = \sum_{n=1}^{N} \psi(x)^T \psi(x_n)$$

$$= \sum_{n=1}^{N} \sum_{i=1}^{M} \psi_i(x) \psi_i(x_n)$$

$$\boxed{\sum_{n=1}^{N} k(x, x_n) = \sum \psi_i(x) \delta_{i1} = \psi_1(x) = 1}$$

3.21

The eigenvector equation for the $M \times M$ real, symmetric A can be written as

$$A u_i = \eta_i u_i$$

$\{u_i\}$ set of M orthonormal vectors, and the M eigenvalues $\{\eta_i\}$

$$\ln |A| = \ln \prod_{i=1}^{M} \eta_i = \sum_{i=1}^{M} \ln \eta_i$$

$$\frac{d}{d\alpha} \ln(|A|) = \sum_{i=1}^{M} \frac{1}{\eta_i} \frac{d}{d\alpha} \eta_i$$

From (3.127),

$$A = \sum_{i=1}^{M} \eta_i u_i u_i^T$$

(3)

$$A^{-1} = \sum_{i=1}^{M} \frac{1}{n_i} u_i u_i^T$$

$$\Rightarrow \mathrm{Tr}\left(A^{-1} \frac{d}{d\alpha} A\right) = \mathrm{Tr}\left(\sum_{i=1}^{M} \frac{1}{n_i} u_i u_i^T \frac{d}{d\alpha} \sum_{j=1}^{M} n_j u_j u_j^T\right)$$

$$= \mathrm{Tr}\left(\sum_{i=1}^{M} \frac{1}{n_i} u_i u_i^T \left\{ \sum_{j=1}^{M} \frac{dn_j}{d\alpha} u_j u_j^T + n_j (b_j u_j^T + u_j b_j^T) \right\}\right)$$

$$= \mathrm{Tr}\left(\sum_{i=1}^{M} \frac{1}{n_i} u_i u_i^T \sum_{j=1}^{M} \frac{dn_j}{d\alpha} u_j u_j^T\right)$$

(244) $$+ \mathrm{Tr}\left(\sum_{i=1}^{M} \frac{1}{n_j} u_i u_i^T \sum^{M} n_j (b_j u_j^T + u_j b_j^T)\right)$$

$$b_j = \frac{du_j}{d\alpha}$$

$$\mathrm{Tr}\left(\sum_{i=1}^{M} \frac{1}{n_i} u_i u_i^T \sum_{j=1}^{M} n_j (b_j u_j^T + u_j b_j^T)\right)$$

$$= \mathrm{Tr}\left(\sum_{i=1}^{M} \frac{1}{n_i} u_i u_i^T \sum_{j=1}^{M} 2 n_j u_j b_j^T\right)$$

$$= \mathrm{Tr}\left(\sum_{i=1}^{M} b_j u_j^T + u_j b_j^T\right)$$

$$= \mathrm{Tr}\left(\frac{d}{d\alpha} \sum_{i}^{M} u_i u_i^T\right)$$

$\bowtie$ and $\sum_{i}^{M} u_i u_i^T = I$

(144)

$$Tr\left(A^{-1}\frac{d}{d\alpha}A\right) = \sum_{i=1}^{M} \frac{1}{\eta_i}\frac{d\eta_i}{d\alpha}$$

$$\frac{d}{d\alpha}\ln p(t|\alpha\beta) = \frac{M}{2}\frac{1}{\alpha} - \frac{1}{2}m_N^T m_N - \frac{1}{2}Tr\left(A^{-1}\frac{d}{d\alpha}A\right)$$

$$= \frac{1}{2}\left(\frac{M}{\alpha} - m_N^T m_N - Tr(A^{-1})\right)$$

$$\boxed{\frac{d}{d\alpha}\ln p(t|\alpha\vec{B})\ \frac{1}{2}\left(\frac{M}{\alpha} - m_N^T m_N - \sum_i \frac{1}{\lambda_i+\alpha}\right)}$$

2. Programming (5)

(a) My experiment is really simple. It consisted on finding the variable that combined FTP and WE gives the smallest cost function.

Formula and Math steps used to get the ~~get~~ cost function

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon.$$

$$\Rightarrow \quad y = X\beta + \varepsilon.$$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & & x_{nn} \end{bmatrix}$

$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$ and $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$$\hat{\beta} = (X'X^{-1})X'y.$$

$$\hat{y} = X\hat{\beta}$$

cost function $= \dfrac{1}{2m} \sum_{i=1}^{m} (y - y_i)^2$

and then Find all the cost functions for all the different possibilities

AFter experiment the variable with smallest cost function is. LIC

using $\hat{B} = (X'X)^{-1}X'y$. we get for LIC

$B_0 = -58.1244$ $B_1 = 0.1847$ $B_2 = 0.1068$

$B_3 = 0.0165$.

The final equation is

$$\boxed{y = -58.1244 + 0.1847 * FTP + 0.1068 * WE + 0.0165 * LIC}$$