I. Handwriting (30%)

1) Consider a Gaussian process regression model in which the target variable t has dimensionality D. Write down the conditional distribution of $t_N + 1$ for a test input vector $x_N + 1$, given a training set of input vectors $x_1 , \dots , x_N + 1$ and corresponding target observations $t_1 , \dots , t_N$.

Ans:

We should evaluate the predictive distribution $p(t_{N+1}|t_N)$ which is also conditioned on$x_1 , \dots , x_N$ and$x_N + 1$ which are not shown to keep notation simple

To find conditional distribution $p(t_{N+1}|t_N)$ we begin by writing down the joint distribution $p(t_{N+1})$

Joint distribution is given by

$$p(t_{N+1}) = N(t_{N+1}|0, C_{N+1})$$

where C_{N+1} is the $(N+1) \times (N+1)$ covariance matrix with elements given by $C(x_n, x_m) = k(x_n, x_m) + \beta^{-1} d_{nm}$

Because the joint is Gaussian, conditional Gaussian distribution is given by partitioning the covariance matrix as

$$C_{N+1} = \begin{pmatrix} C_N & k \\ k^T & c \end{pmatrix}$$

where C_N is the $N \times N$ covariance matrix vector k has elements $k(x_N, x_{N+1})$ for $n = 1, \dots, N$ and scalar $c = k(x_N, x_{N+1}) + \beta^{-1}$

Conditional distribution p(tN+1|t) is Gaussian with

$$\mathit{Mean}: m(x_{N+1}) = k^T C_N^{-1} t$$

$$\mathit{Variance}: \sigma^2(x_{N+1}) = c - k^T C_N^{-1} k$$

Key results that define Gaussian Regression

2) Drive the result

$$\ln p\,(t|X, \alpha, \beta) = \ln N(t|0, C)$$

$$= -\frac{1}{2}\{N \ln(2\pi) + \ln|C| + t^T C^{-1} t\}$$

For the marginal likelihood function in the regression RVM, by performing the Gaussian integral over w in

$$p(t|X, \alpha, \beta) = \int p(t|X, \alpha, \beta) p(w|\alpha) dw$$

Using the technique of completing the square in the exponential.

Ans:

To derive the marginal likelihood function in the regression Relevance Vector Machine (RVM), we start with the joint distribution of the target variable t and the weight vector w:

$$p(t, w|X, \alpha, \beta) = p(t|w, X, \beta)p(w|\alpha)$$

where X is the input data, α and β are the hyperparameters, and $p(t|w, X, \beta)$ is the likelihood function. Assuming a Gaussian likelihood, we have:

$$p(t|w, X, \beta) = N(t|Xw, \beta^{-1})$$

Now we integrate out the weight vector w to obtain the marginal likelihood:

$$p(t, w|X, \alpha, \beta) = \int p(t|w, X, \beta)p(w|\alpha)dw$$

Since both the prior and the likelihood are Gaussian, the posterior distribution $p(w|t, X, \alpha, \beta)$ is also Gaussian. Using the Bayes' theorem, we have:

$$p(w|t, X, \alpha, \beta) \propto p(t|w, X, \beta)p(w|\alpha)$$

Taking the logarithm of both sides, we get:

$$\ln p(w|t, X, \alpha, \beta) - \ln p(t|w, X, \beta) + \ln p(w|\alpha) + constant$$

Expanding the terms, we have:

$$\ln p(w|t, X, \alpha, \beta) = -\frac{\beta}{2}(t - Xw)^T(t - Xw) - \frac{\alpha}{2}w^T w + constant$$

Completing the square in the exponent, we get:

$$\ln p(w|t, X, \alpha, \beta) = -\frac{1}{2}(w - m)^T A(w - m) + constant$$

where $m = \beta A^{-1}X^T t, A = \alpha I + \beta X^T X$

Now we can perform the Gaussian integral over w to obtain the marginal likelihood:

$$p(t, w|X, \alpha, \beta) = \int p(t|w, X, \beta)p(w|\alpha)dw$$

$$= \int N(t|Xw, \beta^{-1})N(w|0, A^{-1})dw$$

$$= N(t|0, C)$$

where $C = \beta^{-1} + XA^{-1}X^T$ is the covariance matrix of the target variable t, and we have used the property of the multivariate normal distribution that the integral of a product of two normal distributions

is also a normal distribution with mean and covariance given by the sum and difference of the means and covariances of the two distributions, respectively.

Taking the logarithm of both sides and simplifying, we get:

$$\ln p(t|X,\alpha,\beta) = -\frac{1}{2}\{N\ln(2\pi) + \ln|C| + t^T C^{-1} t\}$$

which is the desired result.

3) Show that there are $2^{M(M-1)/2}$ distinct indirected graphs over a set of M distinct random variables. Draw the 8 possibilities of the case of $M = 3$.

Ans:

要顯示在一組 M 個不同的隨機變量中，存在$2^{M(M-1)/2}$個不同的有向圖，我們可以使用以下事實：每對不同的變量可以通過有向邊連接或不連接，每對變量有 2 種可能性。由於有 M(M-1)/2 對不同的變量，因此有向圖的總數為$2^{M(M-1)/2}$。

對於 M = 3，我們有 3(3-1)/2 = 3 對不同的變量，因此有$2^{3(3-1)/2} = 2^3 = 8$個不同的有向圖。我們可以將這 8 種可能性繪製如下：

1.

2.

3.

4.

5.

6.

7.

8.

其中一些圖是等價的，即它們表示相同的條件獨立關係。例如，圖1和圖3是等價的，圖2和圖8，圖4和圖6，圖5和圖7也是等價的。這是因為每對變量都是在兩個圖中連接還是沒有連接，因此條件獨立關係是相同的。因此，這8個圖實際上只代表4個不同的條件獨立關係。

4) In Section 7.2.1 we used direct maximization of the marginal likelihood function to derive the re-estimation equations

$$\alpha_i^{new} = \frac{\gamma_i}{m_i^2}$$

$$(\beta^{new})^{-1} = \frac{\|t - \phi m||^2}{N - \sum_i \gamma^i}$$

for finding values of the hyperparamters α and β for the regression RVM. Similarly, in Section 9.3.4 we used the EM algorithm to maximize the same marginal likelihood, giving the re-estimation equations

$$\alpha_i^{new} = \frac{1}{m_i^2 + \sum ii}$$

$$(\beta^{new})^{-1} = \frac{\|t - \phi m_N||^2 + \beta^{-1} \sum_i \gamma_i}{N}$$

Show that these two sets of re-estimation equations are formally equivalent

Ans:

要證明這兩組重新估計方程式在形式上是等價的，我們從以 α 和 β 為參數的邊際概似的表達式開始，這些表達式是:

$$p(t|\alpha,\beta) = \int p(t|w,\beta)p(w|\alpha)dw$$

$$\ln p(t|\alpha,\beta) = \frac{1}{2}\ln|\beta| - \frac{1}{2}(t-\phi m)^T\beta(t-\phi m) - \frac{1}{2}\sum_i(\ln\alpha_i) + \frac{1}{2}\sum_i \alpha_i m_i^2$$

其中 m 是權重向量，γ_i是與第 i 個權重相關聯的精度參數。

對α_i進行直接最大化邊際概似的結果是：

$$\alpha_i^{new} = \frac{\gamma_i}{m_i^2}$$

將此表達式代入邊際概似的表達式中，我們得到：

$$\ln p(t|\alpha_i^{new},\beta) =$$

$$\frac{1}{2}\ln|\beta| - \frac{1}{2}(t-\phi m)^T\beta(t-\phi m) - \frac{1}{2}\sum_i \ln\left(\frac{\gamma_i}{m_i^2}\right) + \frac{1}{2}\sum_i \frac{\gamma_i}{m_i^2}$$

$$= \frac{1}{2}\ln|\beta| - \frac{1}{2}(t-\phi m)^T\beta(t-\phi m) - \frac{1}{2}\sum_i(\ln\gamma_i) + \frac{1}{2}\sum_i \frac{\gamma_i}{m_i^2} - \frac{1}{2}\sum_i \ln(m_i^2)$$

對此表達式關於β進行求導並設置為零，得到：

$$(\beta^{new})^{-1} = \frac{||t-\phi m||^2}{N - \sum_i \gamma^i}$$

同樣地，對於 EM 算法，我們從期望對 w 的後驗分布取期望值的表達式開始，該表達式是：

$$E[ln\, p(t,w|\alpha,\beta)] = E[ln\, p(t|w,\beta)] + E[ln\, p(w|\alpha)]$$

對此表達式關於α_i進行求導並設置為零，得到：

$$\alpha_i^{new} = \frac{1}{m_i^2 + \sum ii}$$

將此表達式代入期望對 w 的後驗分布取期望值的表達式中，我們得到：

$$E[\ln p(t, w|\alpha_i^{new}, \beta)] = E[\ln p(t|w, \beta)] + E[\ln p(w|\alpha_i^{new})]$$

對此表達式關於β進行求導並設置為零，得到：

$$(\beta^{new})^{-1} = \frac{||t - \phi m_N||^2 + \beta^{-1} \sum_i \gamma_i}{N}$$

其中m_N是 w 的後驗分布的平均值，γ_i是第 i 個權重相關的精度參數。

因此，我們已經證明了這兩組重新估計方程在形式上是等價的。