问题 1：

问题 1 第一小问

推导过程(涉及高斯分布，贝叶斯公式，分层贝叶斯)

该问中基于传统方法(譬如最小二乘)这些方法无法考虑观测值过程中的存在的一些观测噪声，因此，在这里我们引入贝叶斯推导框架，将观测过程中存在的观测噪声考虑在内(指下文公式中的 E)；总的来说，该问题可以转化为回归方程问题求解，即存在观测方程为

$$Y = XG + E \tag{1}$$

其中，Y 为观测值，$Y = \left[y_1, y_2\right], Y \in R^{Tem\times 2}, y_i = \left[y_{i,1}, y_{i,2}, \cdots, y_{i,Tem-1}, y_{i,Tem}\right]^T \in R^{Tem\times 1}$，G 是传递矩阵（即是我们需要计算解答的），$G \in R^{2\times 2}$，X 是权重输入自变量(即温度)，$X \in R^{Tem\times 2}$；E 是观测噪声 $E \in R^{Tem\times 2}$。

那么,基于式(1)，可以进一步表示成

$$y_k = Xg_k + e_k \tag{2}$$

其中，$y_k \in R^{Tem\times 1}, X \in R^{Tem\times 2}, g_k \in R^{2\times 1}, e_k \in R^{Tem\times 1}$，$\bullet_k$ 表示第 k 列向量。

我们知道，传统的贝叶斯推导公式为：

$$P\left(A \mid B\right) = \frac{P\left(A, B\right)}{P(B)} \tag{3}$$

而在式（2）中，我们需要求解的为矩阵 g_k，那么我们可以基于公式(2)构建关于矩阵 g_k 的后验概率密度，即有

$$p\left(g_k \mid y_k, \theta\right) = \frac{p\left(y_k \mid g_k, \theta\right) p\left(g_k \mid \theta\right)}{p\left(y_k \mid \theta\right)} = \frac{p\left(y_k \mid g_k, \theta\right) p\left(g_k \mid \theta\right)}{\int p\left(y_k \mid g_k, \theta\right) p\left(g_k \mid \theta\right) dg} \tag{4}$$

其中，θ 表示超参数集合。

在这里，我们假设 g_k 服从高斯分布，那么有

$$g \sim N\left(0, \Sigma_g\right) \tag{5}$$

其中

$$\Sigma_g = diag\left\{\alpha_1, \alpha_2, \cdots, \alpha_M\right\} \tag{6}$$

那么关于 g_k 的先验概率密度函数服从

$$p\left(g \mid \Sigma_g\right) = \prod_{m=1}^{M} p\left(g_m \mid \alpha_m\right) = \left(2\pi\right)^{-\frac{M}{2}} \left|\Sigma_g\right|^{-\frac{1}{2}} e^{-\frac{1}{2} g^T \Sigma_g^{-1} g} \tag{7}$$

下面我们假设噪声e_k为高斯白噪声，那么有

$$e_k \sim N\left(0, \Sigma_v\right) \tag{8}$$

$$\Sigma_v = \sigma^2 I \tag{9}$$

所以，优于高斯分的性质我们可知

$$e_k = y_k - Xg_k \sim N\left(0, \Sigma_v\right) \tag{10}$$

那么既有

$$y_k \sim N\left(Xg_k, \Sigma_v\right) \tag{11}$$

关于观测值y_k的概率密度函数可以表示为

$$p\left(y_k\right) = \left(2\pi\right)^{-\frac{N}{2}} \left|\Sigma_v\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\left(y_k - Xg_k\right)^T \Sigma_v^{-1}\left(y_k - Xg_k\right)} \tag{12}$$

那么，基于式(4)的g_k的后验概率密度可以表示为

$$p\left(g_k \mid y_k, \theta\right) = N\left(g_k \mid Cy_k, P\right) \tag{13}$$

其中

$$P = \left(X^T \Sigma_v^{-1} X + \Sigma_g^{-1}\right)^{-1} \tag{14}$$

$$C = PX^T \Sigma_v^{-1} \tag{15}$$

公式求解过程(设计变分贝叶斯，EM 算法，KL 熵)

接下来，为了得到一个合适的值，并对公式中相应参数的值进行求解，我们在这里引入变分贝叶斯和 EM 算法来计算边际似然概率密度函数的大小，并使得边际似然概率密度函数达到最大。那么，结合 EM 算法，我们可以得到一个关于 y 边际似然的目标函数

$$Q\left(\theta, \widehat{\theta}^{(k)}\right) = E_{p\left(g \mid y, \theta^{(k)}\right)}\left[\ln p\left(y \mid g, \theta\right) + \ln p\left(g \mid \theta\right)\right] \tag{16}$$

并通过最大化该目标函数求得参数值

$$\widehat{\theta}^{(k+1)} = \underset{\theta}{\arg\max}\ Q\left(\theta, \widehat{\theta}^{(k)}\right) \tag{17}$$

目标函数可分为 A,B,C 三部分,即

$$Q\left(\theta, \widehat{\theta}^{(k)}\right) = \underbrace{E_{p\left(g_k \mid y_k, \theta^{(k)}\right)} \ln p\left(y_k \mid g_k, \theta\right)}_{A} + \underbrace{E_{p\left(g_k \mid y_k, \theta^{(k)}\right)} \ln p\left(g_k \mid \theta\right)}_{B} \tag{18}$$

下面对 A,B 分别求积分，即有

A Section

$$E_{p\left(g_k \mid y_k, \theta^{(k)}\right)} \ln p\left(y_k \mid g_k, \theta\right) = -\frac{1}{2}\{N \ln 2\pi + \ln\left|\Sigma_v\right| + y^T \Sigma_v^{-1} y - 2y_k^T \Sigma_v^{-1} UC^{(k)} y_k \tag{19}$$
$$+ trace\left\{\left(X^T \Sigma_v^{-1} X\right) P^{(k)}\right\} + y_k^T \left(C^{(k)}\right)^T \left(X^T \Sigma_v^{-1} X\right) C^{(k)} y_k\}$$

B Section

$$E_{p\left(g_k|y_k,\theta^{(k)}\right)} \ln p\left(g_k \mid \theta\right) = -\frac{1}{2}\left\{M \ln 2\pi + \ln\left|\Sigma_g\right| + trace\left\{\Sigma_g^{-1} P^{(k)}\right\} + y_k^T \left(C^{(k)}\right)^T \Sigma_g^{-1} C^{(k)} y_k\right\} \tag{20}$$

于是

$$\begin{aligned} Q\left(\theta,\widehat{\theta}^{(k)}\right) &= -\frac{1}{2}\{N \ln 2\pi + \ln\left|\Sigma_v\right| + y_k^T \Sigma_v^{-1} y_k - 2 y_k^T \Sigma_v^{-1} X C^{(k)} y_k + trace\left\{\left(X^T \Sigma_v^{-1} X\right) P^{(k)}\right\} \\ &+ y_k^T \left(C^{(k)}\right)^T \left(X^T \Sigma_v^{-1} X\right) C^{(k)} y_k + M \ln 2\pi + \ln\left|\Sigma_g\right| + trace\left\{\Sigma_g^{-1} P^{(k)}\right\} + y_k^T \left(C^{(k)}\right)^T \Sigma_g^{-1} C^{(k)} y_k\} \end{aligned} \tag{21}$$

Section 3 Calculate σ^2

$$\begin{aligned} Q\left(\sigma^2,\widehat{\theta}^{(k)}\right) &= -\frac{1}{2}\{\ln\left|\Sigma_v\right| + y_k^T \Sigma_v^{-1} y_k - 2 y_k^T \Sigma_v^{-1} U C^{(k)} y_k \\ &+ trace\left\{\left(X^T \Sigma_v^{-1} X\right) P^{(k)}\right\} + y_k^T \left(C^{(k)}\right)^T \left(X^T \Sigma_v^{-1} X\right) C^{(k)} y_k\} \end{aligned} \tag{22}$$

$$\frac{dQ\left(\sigma^2,\widehat{\theta}^{(k)}\right)}{d\sigma^2} = 0 \tag{23}$$

$$\left(\sigma^2\right)^{(k+1)} = \frac{y_k^T y_k + \sum S(t,t) + y_k^T C^T X^T X C y_k - 2 y_k^T U C y_k}{N} \tag{24}$$

$$S^{old} = X P^{old} X^T \tag{25}$$

Section 4 Calculate Σ_g

$$Q\left(\alpha_m,\widehat{\theta}^{(k)}\right) = -\frac{1}{2}\left\{\ln\left|\Sigma_g\right| + trace\left\{\Sigma_g^{-1} P^{(k)}\right\} + y_k^T \left(C^{(k)}\right)^T \Sigma_g^{-1} C^{(k)} y_k\right\} \tag{26}$$

$$\frac{dQ\left(\alpha_m,\widehat{\theta}^{(k)}\right)}{d\alpha_m} = 0 \tag{27}$$

$$\alpha_m = P^{(k)}\left(m,m\right) + \left(O^{(k)}\left(m\right)\right)^2 \tag{28}$$

Where

$$O^{(k)}\left(m\right) = C^{(k)} y_k \tag{29}$$

$$\Sigma_g = diag\left\{\alpha_1^{(k+1)}, \alpha_2^{(k+1)}, \cdots, \alpha_M^{(k+1)}\right\} \tag{30}$$

那么，基于以上总的来说，更新公式可以表述为如下

$Input: \theta^{Ini}, y_k, X$

$and\ we\ can\ define\ that\ \theta^{Ini} = \left[\sigma_{ini}^2, \alpha_1, \alpha_2, \cdots, \alpha_M\right].$

$in\ addition, y_k\ is\ the\ t\arg et\ value.$

$And\ U\ is\ the\ transform\ matrix\left(we\ can\ \exp ress\ X\ as\ 'regressors'\ \text{in}\ this\ system\right)$

$Output: g_k$

$$Step1: repeat\ until\ \frac{\left\|Q\left(\theta, \widehat{\theta}^{(k+1)}\right) - Q\left(\theta, \widehat{\theta}^{(k)}\right)\right\|}{\left\|Q\left(\theta, \widehat{\theta}^{(k)}\right)\right\|} < \varepsilon, \varepsilon\ is\ any\ small\ \text{number}$$

$Step1-1: Calculate\ \sigma^2 \text{by}\ (24)$

$Step1-2: Calculate\ \alpha_m \text{by}\ (28)\ \text{and update}\ \Sigma_g by(30)$

$Step2: Compute\ g_k = C^{(k)} y_k$

推导公式已经全部完成。(上面一字不改的要照抄，可以改一些语句通顺的地方，但是涉及公式和参数的地方千万不要改)

该方法的优点：

引入了贝叶斯框架，即相应参数的先验分布，这使得我们可以在一定程度上提高参数估计的准确率。同时，引入 EM 算法和变分贝叶斯，这使得我们的解为全局最优解，而不是局部最优；不仅如此，基于贝叶斯发展的回归方程求解模型，可以在一定程度上压制噪声，相对于传统方法，在参数估计上具有一定的性能提升。

第一题中问题要求我们探究**乙醇转化率**和 **C4 烯烃选择性**与**温度**的关系，已知附件一中共有 **21** 种不同的催化剂，那么便可以得到 21 个 A 矩阵，即构成了 21 个不同的回归方程**(这里需要手动替换 Y 和 X 中的数据)**。通过该方程，来说明乙醇转化率和 C4 烯烃选择性与温度的关系。在这里，公式(1)构成了一个通用的回归方程矩阵。其中不同的是传递矩阵 A 的变化来说明不同催化剂下温度和其他指标的关系。

（结果展示，有 21 个催化剂，那么可以得到 21 个矩阵 A 构建的 21 个回归方程，我们可以将这 21 个方程列出来。建议论文中只列出一个，剩下 20 个放进附录）

运行 Optimized_Question_1_1.m 文件

另一个.m 文件千万不要动