Master de Chemoinformatique et Modélisation M1S1
Examen de mathématiques pour la chimie*

MARCOU Gilles

2012-2013

Résumé

Documents autorisés. Durée : 2h. La copie de l'étudiant sera constituée d'un support papier traditionnel et d'un ou plusieurs fichiers de calcul pour Maple. ***Toute réponse doit être motivée ou sera considérée comme nulle.***

Exercice 1

In this excercise, you will build a statistical model in order to estimate the stability constant (pKHB) of complexes stablized by one hydorgen bond. The study will be limited to a serie of 6-membered niytrogen-heterocyclic compounds. The study contains $N = 54$ compounds and each compound is characterized by 4 molecular descriptors derived from AM1 semiempirical molecular orbitals (Coulson charge on the nitrogen, Eigenvalue of the localized nitrogen lone pair orbital, Coefficient of the p-orbital in the nitrogen lone pair orbital, Access angle) [J Mol Model (2002) 8 :95-101]. This exercise illustrates the PLS method.

The dataset is available as an Excel sheet (`pKHBdata.xls`). For this exercise, the molecular descriptors have been stored in a space separated data file (`X.txt`) and the experimental pKHB to model are stored in separate file (`Y.txt`). Those data are loaded in memory in this Maple worksheet, into the matrix $[X]$ and the column vector $|Y\rangle$. Each column of $[X]$ corresponds to one molecular descriptor. The property and the molecular descriptors are standardized.

Question 1.1

Using the function **Covariance**, from the **Statistics** package, compute the covariance between each of the d columns of $[X]$ and the property $|Y\rangle$. Store the results in a vector named $|t_1\rangle$ and normalize $|t_1\rangle$ using the command **Normalize**.

Question 1.2

Compute the hat matrix

$$[H_1] = [X]|t_1\rangle(\langle t_1|^t[X][X]|t_1\rangle)^{-1}\langle t_1|^t[X] \quad (1)$$

and the orthogonal projection matrix $[P_1] = 1 - [H_1]$.

Applying the projection matrix $[P_1]$ to the property vector give the error of the fit of this partial model. Beside, applying the same projection matrix $[P_1]$ to the original factor matrix $[X]$ give a new set of molecular descriptors that are orthogonal to the one selected at this initial stage. Applying the hat matrix $[H_1]$ to the vector $|Y\rangle$ give the current state of the partial fit.

*Enseignants: G. Marcou, *Université de Strasbourg, Faculté de Chimie, 1, rue Blaise Pascal, 67000 Strasbourg*

Question 1.3

Compute the error of the partial fit, $|E_1\rangle = [P_1]|Y\rangle$ and the next factor matrix $[PX_1] = [P_1][X]$ and the fitted values $|Y_{\text{pred}}\rangle = [H_1]|Y\rangle$.

Question 1.4

Compute the vector of covariance $|t_2\rangle$ between $[PX_1]$ and the unexplained part of property $[E_1]$ and normalize it.

Question 1.5

Compute the hat matrix

$$[H_2] = [PX_1]|t_2\rangle(\langle t_2|^t[PX_1][PX_1]|t_2\rangle)^{-1}\langle t_2|^t[PX_1] \tag{2}$$

and the related projection matrix $[P_2] = 1 - [H_2]$.

The current state of the fit is simply $[H_1]|Y\rangle + [H_2]|E_1\rangle$. The current residues are obtained using the projection matrix $|E_2\rangle = [P_2]|E_1\rangle$. The procedure can be iterated until the number of original descriptors is reached. At that point the PLS is equivalent to a classical multi-linear regression.

Question 1.6

Resume the previous steps to write a PLS function.

Exercice 2

This exercise proposes to model titration curves. Such titration curve relates the pH of an acido-basic solution with the volume of titrant. This well known curve, however, has no explicit formulation, unless some approximations are used regarding the relative strength of the acid and base. In order to fix ideas, but without loss of generalization, the exercise concerns the titration of a monoprotic acid with a monoprotic base.

Note that some variables are initialized with mathematical symbols and relations in order to simplify the manipulation of expressions.

The system is governed by three equilibrium constants :

$$Ka_1 = \frac{A^- H_3O^+}{AH} \tag{3}$$

$$Ka_2 = \frac{BH_3O^+}{BH^+} \tag{4}$$

$$Ke = H_3O^+OH^- \tag{5}$$

A^-, AH, BH^+, B, H_3O^+, OH^- are respectively the concentrations of ionized and neutral acid, ionized and neutral base, hydronium and hydroxide ions. Beside, the system is electrostatically neutral.

Question 2.1

Find an expression of the fraction of dissociated acid $\alpha = \frac{A^-}{C_A}$ and of protonated base $\beta = \frac{BH^+}{C_B}$ where C_A and C_B are the analytical concentrations of acid and base respectively. These expressions shall involve only the equilibirium constants Ka_1 and Ka_2 and the hydronium concentration H_3O^+. The following lines should answer to this question, but they are broken : the command **subs** should be replaced by a more relevant one.

Question 2.2

Reformulate the electrostatic equation so that it includes solely, hydronium concentration, the analytical concentrations, and the fractions α and β.

Question 2.3

Give an expression of the analytical concentrations C_A and C_B using the molarity and volumes of the acid (ρ_A, V_A) and of the base (ρ_B, V_{Tr}). The titrant being the base, the volume of the titrant V_{Tr}.

Question 2.4

Reformulate the electrostatic equation so that it contains only references to the molarities, volumes, analytical concentrations and hydronium concentration.

Note that this equation is non-linear in terms of the concentration of hydronium and there is no analytical expression of this concentration as a function of the volume of titrant. However, it is much more easy to find an analytical expression of the volume of titrant as a function the concentration of hydronium.

Question 2.5

Solve the electrostatic equation using the volume of titrant $V_{Tr} = y$ as unknown and replacing $H_3O^+ = 10^{-x}$. In order to simplify the manipulation of this equation, some change of notations were operated. Plot the solution after setting reasonable values for the parameters of the problem.

Barème

- Question 1.1 : 2 points
- Question 1.2 : 2 points
- Question 1.3 : 2 points
- Question 1.4 : 2 points
- Question 1.5 : 2 points
- Question 1.6 : 2 points
- Question 2.1 : 2 points
- Question 2.2 : 2 points
- Question 2.3 : 2 points
- Question 2.4 : 2 points
- Question 2.5 : 2 points