Τρόπος επεξεργασίας δεδομένων χαμηλής αβεβαιότητας

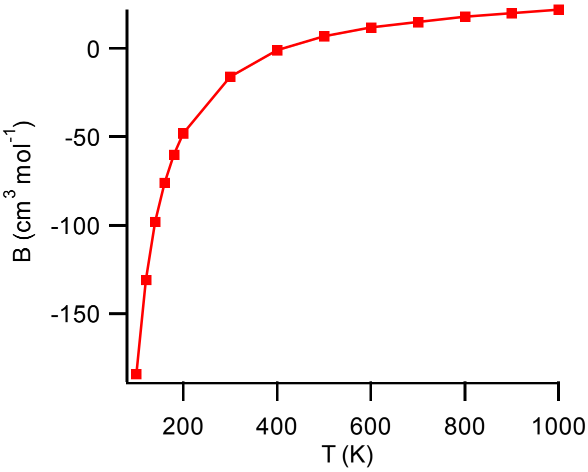
Είναι συνηθισμένο φαινόμενο η υποβάθμιση της ποιότητας σειράς δεδομένων λόγω εσφαλμένης επεξεργασίας. Η συνηθισμένη επεξεργασία είναι υπολογισμός ευθείας με τη μέθοδο των ελαχίστων τετραγώνων. Αυτή προϋποθέτει γραμμική σχέση μεταξύ των ποσοτήτων, πράγμα που δεν είναι προφανές και πολλές φορές διαψεύδεται από τα δεδομένα.

Παράδειγμα.

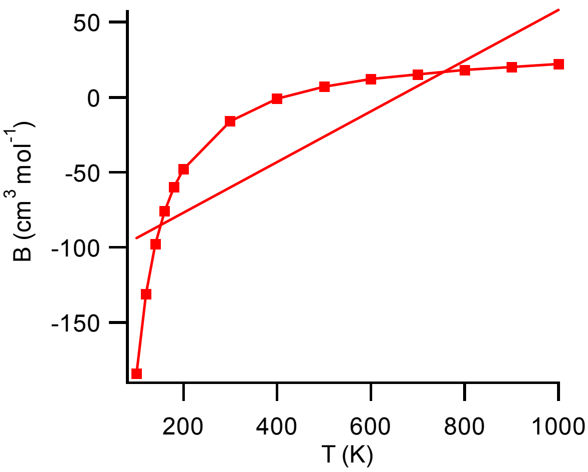
Δίνονται τιμές παρμένες από την βιβλιογραφία για τον δεύτερο συντελεστή virial για το αέριο αργό (Ar) συναρτήσει θερμοκρασίας και χρειαζόμαστε την τιμή του συντελεστή για θ = 22°C.

T (K)	B (cm^3 mol^{-1})
100	-184
120	-131
140	-98
160	-76
180	-60
200	-48
300	-16
400	-1
500	7
600	12
700	15
800	18
900	20
1000	22

Σχεδιάζουμε τα δεδομένα για να αποφασίσουμε πώς θα υπολογίσουμε την ζητούμενη τιμή σε θερμοκρασία T = 295 K.

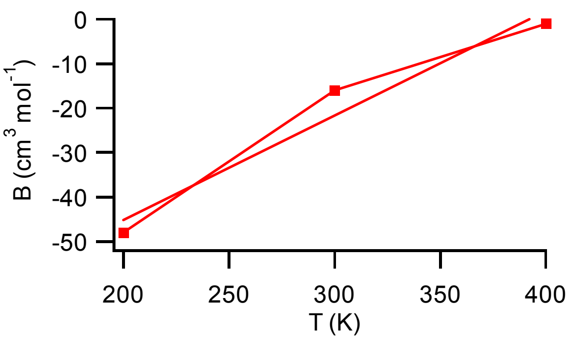


Μία συνηθισμένη τακτική είναι να υπολογίσουμε με «ελάχιστα τετράγωνα» μια ευθεία και με την εξίσωσή της να υπολογίσουμε το B για T = 295 K. Να το σχετικό αποτέλεσμα:



Η εξίσωση της ευθείας είναι B = -110.63 cm^3 mol^{-1} +0.16865 cm^3 mol^{-1} K^{-1} × T, απ’ όπου προκύπτει ότι B = -60.9 cm^3 mol^{-1}.

Αν αγνοήσουμε τα σημεία με T < 200 K και T > 400 K, έχουμε το διάγραμμα:



Αν χρησιμοποιήσουμε πάλι την μέθοδο ελαχίστων τετράγωνων, η ευθεία που προκύπτει έχει εξίσωση B = -92.167 cm^3 mol^{-1} +0.235 cm^3 mol^{-1} K^{-1} × T, απ’ όπου προκύπτει ότι B = -22.8 cm^3 mol^{-1}. Αν χρειαζόμασταν την τιμή για T = 300 K, η τιμή που υπολογίζουμε είναι -21.7 cm^3 mol^{-1} αντί του σωστού (από τον πίνακα) -16 cm^3 mol^{-1}. Προφανώς ούτε αυτή η διαδικασία δίνει ικανοποιητικό αποτέλεσμα.

Αν ξεχάσουμε την μέθοδο ελαχίστων τετραγώνων και βασιστούμε μόνο στα πλησιέστερα δεδομένα, δηλ. τις τιμές του B για T = 200 K και T = 300 K. Μπορούμε να υπολογίσουμε γενική εξίσωση για κάθε σημείο του διαστήματος (200, 300) με κλίση

$$\alpha = \frac{y_2 - y_1}{x_2 - x_1} = \frac{B_2 - B_1}{T_2 - T_1} = \frac{-16-(-48)}{300-200}\frac{\mathrm{cm}^3}{\mathrm{mol\,K}} = 0.32\frac{\mathrm{cm}^3}{\mathrm{mol\,K}}$$

και σταθερό όρο

$$\beta = y_1 - \alpha x_1 = -48\frac{\mathrm{cm}^3}{\mathrm{mol}} - 0.32\frac{\mathrm{cm}^3}{\mathrm{mol\,K}} \times 200\,\mathrm{K} = -112\frac{\mathrm{cm}^3}{\mathrm{mol}},$$

οπότε B(295K) = -112 $\mathrm{cm^3\,mol^{-1}}$+0.32 $\mathrm{cm^3\,mol^{-1}\,K^{-1}}$ × 295 K = -17.6 $\mathrm{cm^3\,mol^{-1}}$.

Μπορούμε να κάνουμε τον ίδιο υπολογισμό ως διόρθωση στην πλησιέστερη προς την επιθυμητή τιμή: αφού υπολογίσουμε την κλίση, διορθώνουμε την τιμή της T = 300 K κατά ποσό ανάλογο της απόστασή της από την T = 295 K, δηλ. B(295 K) = B(300 K) – α (300 – 295) K = -16 -0.32 × 5 = -17.6 $\mathrm{cm^3\,mol^{-1}}$.

Αυτός ο απλός υπολογισμός είναι ο ακριβέστερες από τους τρεις που αναφέραμε. Οι άλλοι δύο είναι λανθασμένοι διότι κάνουν εσφαλμένη υπόθεση για γραμμική συσχέτιση μεταξύ T και B. Το ότι δεν ταιριάζει καλά μια ευθεία στις δύο πρώτες απόπειρες δεν οφείλεται στα σφάλματα των τιμών του διαγράμματος, αλλά σε ακατάλληλη συναρτησιακή μορφή της B = f(T). Η συνάρτηση η οποία περιγράφει ακριβώς όλα τα δεδομένα του διαγράμματος έχει τη μορφή: B = [-16 -60 × (T_0/T-1) -10 × (T_0/T-1)^2] $\mathrm{cm^3\,mol^{-1}}$. Τότε η ζητούμενη τιμή του δεύτερου συντελεστή virial για το αργό σε T = 295 K είναι B = 16.6 $\mathrm{cm^3\,mol^{-1}}$.