

Tracce Prova Orale

Scheda n. 1

1. Nell'ambito delle relazioni quantitative struttura-attività (QSAR), l'esplorazione multivariata mediante Analisi delle Componenti Principali (PCA) richiede un percorso strutturato per estrarre l'informazione latente dai descrittori molecolari. Il/la candidato/a descriva nel dettaglio le diverse fasi operative per il calcolo e l'analisi di un modello PCA, partendo dalle tecniche di pretrattamento dei dati, passando per la scelta del numero di Componenti Principali, fino all'impiego di Score Plot e Loading Plot.
2. Descrivere la procedura operativa per il calcolo di un modello QSAR (Quantitative Structure-Activity Relationship) per la predizione quantitativa della tossicità acquatica acuta verso la *Daphnia Magna* in ambiente R o Python. Descrivere, anche con l'ausilio del dataset-esempio toxicity.xlsx messo a disposizione, le diverse operazioni necessarie per la calibrazione e validazione di un modello di regressione, con particolare attenzione agli aspetti di predittività, interpretabilità degli effetti delle variabili, presenza di dati anomali. Come risposta (quantitativa) del modello si utilizzino i dati di LC50, ovvero la concentrazione che causa la morte nel 50% delle *Daphnia Magna* sottoposte a test nell'arco di una durata di 48 ore. Il dataset-esempio (toxicity.xlsx) include 545 sostanze organiche, elencate nelle righe di colonna A mediante uno specifico codice alfanumerico. Per ciascuna sostanza, sono stati calcolati 8 descrittori molecolari: TPSA(Tot) (Molecular properties), SAacc (Molecular properties), H-050 (Atom-centred fragments), MLOGP (Molecular properties), RDCHI (Connectivity indices), GATS1p (2D autocorrelations), nN (Constitutional indices), C-040 (Atom-centred fragments). La colonna J contiene i valori di LC50.
3. Utilizzando la postazione messa a disposizione dalla Commissione, il/la candidato/a individui le principali caratteristiche del sistema operativo installato e le risorse hardware disponibili (processore, memoria, spazio disco), illustrando le operazioni eseguite.

Prova di informatica:

Il candidato esegua un'elaborazione dati utilizzando il programma Rstudio ed Excel con il caso d'uso indicato nella domanda 2.

Scheda n. 2

1. Nell'ambito delle relazioni quantitative struttura-attività (QSAR), la predizione qualitativa dell'attività biologica richiede un flusso di lavoro volto a separare correttamente le molecole in classi distinte. Il/la candidato/a descriva le tappe necessarie per lo sviluppo di un modello QSAR di classificazione, dettagliando le fasi di calibrazione e validazione del modello, per poi affrontare la valutazione delle prestazioni in termini di accuratezza e la successiva eventuale interpretazione del ruolo discriminatorio svolto dai singoli descrittori.
2. Descrivere la procedura operativa per l'esplorazione e l'analisi della struttura multidimensionale dei dati chimici relativi alla tossicità acquatica acuta verso la *Daphnia Magna*, mediante l'impiego dell'Analisi delle Componenti Principali (PCA) in ambiente R o Python. Descrivere, anche con l'ausilio del dataset di esempio fornito, le diverse operazioni necessarie per il pre-trattamento e l'esplorazione della struttura dei dati, l'individuazione di eventuali dati anomali, e l'interpretazione dell'effetto delle variabili sulla tossicità. Il dataset di esempio include 545 sostanze organiche, elencate nelle righe della prima colonna mediante uno specifico codice alfanumerico. Per ciascuna sostanza sono stati calcolati otto descrittori molecolari: TPSA(Tot) (Molecular properties), SAacc (Molecular properties), H-050 (Atom-centred fragments), MLOGP (Molecular properties), RDCHI (Connectivity indices), GATS1p (2D autocorrelations), nN (Constitutional indices), C-040 (Atom-centred fragments). La colonna J contiene i valori di LC50. Le molecole sono state suddivise in due classi: la classe 1 include molecole con valori di LC50 inferiori a 4 (classe positiva), la classe 2 include invece molecole con valori di LC50 superiori a 4 (classe negativa). La colonna K contiene le etichette di classe.
3. Utilizzando una macchina virtuale predisposta dalla Commissione, il/la candidato/a ne illustri il funzionamento e mostri come accedere al sistema operativo ospite e verificarne le principali caratteristiche.

Prova di informatica:

Il candidato esegua un'elaborazione dati utilizzando il programma Rstudio ed Excel con il caso d'uso indicato nella domanda 2.