UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

# COURSE SYLLABUS

# Chemometrics

**2122-1-F5401Q018**

## Obiettivi

Obiettivo principale dell'insegnamento è fornire allo studente i fondamenti teorici e gli strumenti operativi delle principali tecniche chemiometriche, necessarie nella chimica moderna per il trattamento adeguato dell'informazione contenuta nei dati chimici sperimentali. L'obiettivo del corso è quindi quello di procurare allo studente gli elementi fondamentali dell'analisi multivariata per trattare sistemi complessi di interesse chimico, farmacologico e ambientale. La conoscenza dei principi e della condizioni operative delle principali tecniche chemiometriche verrà sviluppata insieme alla capacità di scegliere e di gestire gli approcci di indagine più adatti alle finalità dell'analisi. Lo studente saprà quindi valutare le caratteristiche degli approcci chemiometrici fondamentali, i campi di applicazione, i vantaggi e gli svantaggi delle singole tecniche e sarà quindi in grado di suggerire la scelta della tecnica chemiometrica ritenuta più idonea per uno specifico problema.

In particolare, al termine del corso di Chemiometria, lo studente dovrà dimostrare di aver raggiunto i seguenti obiettivi formativi:

**Conoscenza e capacità di comprensione:**

- descrivere i principali metodi chemiometrici presentati nel corso al fine di un utilizzo corretto
- descrivere i parametri fondamentali per la valutazione dei risultati
- descrivere vantaggi e svantaggi dei diversi algoritmi studiati

**Capacità di applicare conoscenza e comprensione:**

- selezionare il metodo più adatto ad affrontare un problema di analisi dei dati
- prevedere quale tipo di informazione sarà possibile enucleare dai dati in esame
- valutare la possibilità di metodi chemiometrici alternativi per la soluzione del problema
- applicare in concreto i metodi chemiometrici selezionati e calcolare i relativi parametri statistici informativi

**Autonomia di giudizio:**

- acquisire le conoscenze e competenze in grado di sviluppare una capacità di comprensione critica dei principali metodi chemiometrici
- selezionare le tecniche ed i parametri statistici utili ad estrarre specifiche informazioni dai dati in esame
- redarre e giustificare una discussione critica sui metodi utilizzati e le informazioni ottenute dall'analisi dei dati

**Capacità di apprendimento:**

- comprendere le diverse tecniche studiate e la loro collocazione metodologica al fine di un impiego corretto e mirato al problema da risolvere

## Contenuti sintetici

Introduzione alla chemiometria. La struttura dei dati multivariati. Strategie per la razionalizzazione di problemi complessi: Analisi delle Componenti Principali. Similarità e diversità. Metodi di Cluster Analysis. Il concetto di bias e i metodi di validazione. Metodi di regressione multivariata. Metodi di classificazione multivariata. Accenno alle reti neurali ed alla selezione delle variabili. Strategie di data fusion. Analisi delle relazioni tra struttura molecolare, proprietà chimico-fisiche e attività biologiche (QSAR).

Esperienze pratiche in laboratorio per acquisire gli strumenti e le modalità di analisi tramite le principali tecniche chemiometriche descritte nel corso (analisi della struttura di dati chimici, realizzazione di modelli di regressione e classificazione).

## Programma esteso

Introduzione alla chemiometria: obiettivi, metodi e applicazioni della chemiometria per l'analisi di sistemi chimici complessi. La struttura dei dati multivariati. Elementi di calcolo matriciale. Parametri statistici elementari: indici di posizione e di dispersione, covarianza e correlazione. Le scalature ed il pretrattamento dei dati: centratura, autoscalatura, scalatura di intervallo, scalatura sulla varianza.

Strategie per la razionalizzazione di problemi complessi, l'analisi della struttura e l'esplorazione dei dati chimici connessi a sistemi complessi; l'analisi delle componenti principali (PCA): obiettivi della PCA, accenni all'algoritmo di diagonalizzazione, matrici degli scores e dei loadings; autovalori e definizione delle componenti significative (analisi di rango). Esempi di applicazione della PCA su dati chimici. La correlazione multivariata.

Analisi della similarità e diversità in sistemi complessi: i concetti di analogia, similarità, dissimilarità e distanza. Misure di distanza e similarità per dati quantitativi e binari. Cluster Analysis: metodi gerarchici agglomerativi e metodi non gerarchici. Strategie per l'analisi di similarità. Esempi di applicazione della Cluster Analysis su dati chimici.

Il concetto di bias e i metodi di validazione: stimatori statistici; bias e varianza. Modelli descrittivi e predittivi. Tecniche di validazione dei modelli statistici multivariati: cross-validation, bootstrap, leave-one-out, leave-many-out, y scrambling.

Metodi di regressione multivariata: strategie di ricerca basate su modelli quantitativi e parametri di regressione.

L'analisi di regressione multipla. I metodi di regressione biased: i metodi ridge, di selezione dei migliori sotto modelli, regressione in componenti principali, metodo Partial Least Square. Gli algoritmi genetici per la selezione di variabili. Il metodo Sequential Replacement. Esempi di applicazione della regressione multivariata su dati chimici.

Metodi di classificazione multivariata: strategie di ricerca basate sulla classificazione e parametri di classificazione. I metodi di classificazione locale: k Nearest Neighbours (kNN), N3, BNN. Le probabilità bayesiane e i metodi di analisi discriminante lineare e quadratica. Metodi di classificazione ad albero (CART). Le reti neurali di Kohonen.

Metodi di consensus e data fusion: introduzione alle strategie moderne per la concatenazione di differenti sorgenti di informazione chimica tramite approcci di consensus analysis e data fusion; definizione dei livelli di data fusion.

Introduzione alle relazioni tra struttura molecolare, proprietà chimico-fisiche, biologiche, ambientali (QSAR): metodologie QSAR, descrittori molecolari e loro applicazione.

Sono previste tre esperienze pratiche su dati reali per acquisire gli strumenti e le modalità di analisi sulle seguenti tematiche chemiometriche: analisi della struttura di dati chimici tramite Analisi delle Componenti Principali, realizzazione di modelli di regressione e classificazione multivariata. Le esperienze pratiche vengono eseguite in laboratorio informatico tramite il software MATLAB e specifici toobox grafici forniti dai docenti. Nelle esperienze è prevista una breve introduzione all’utilizzo del sofwtare MATLAB (import e gestione dei dati, integrazione con i toolbox statistici multivariati forniti dai docenti).

## Prerequisiti

Nozioni di base sui principali indici statistici elementari, capacità operativa informatica di base nelle esperienze pratiche in laboratorio.

## Modalità didattica

Il corso si suddivide in una parte di lezioni ed esercitazioni frontali, in cui vengono fornite le nozioni teoriche sulle tematiche chemiometriche. Alla fine dello svolgimento delle lezioni del corso, gli studenti seguono tre differenti esperienze pratiche in laboratorio per acquisire gli strumenti e le modalità operative di analisi delle principali tecniche chemiometriche.

## Materiale didattico

Il libro di testo di riferimento del corso è il seguente: R.Todeschini, Introduzione alla Chemiometria (Edises, Napoli 1998). Il libro viene fornito anche in formato pdf sulla pagina e-learning del corso. Vengono fornite inoltre le slide delle lezioni sulla pagina e-learning del corso. Inoltre, i docenti forniscono tramite piattaforma e-learning un file chm comprendente articoli scientifici per l’approfondimento degli argomenti presentati nel corso. Per ogni esperienza di laboratorio, tramite piattaforma e-learning, sono fornite le slide introduttive, i dati e i toolbox necessari per lo svolgimento delle esperienze tramite software MATLAB.

## Periodo di erogazione dell'insegnamento

Primo semestre

## Modalità di verifica del profitto e valutazione

L'esame consiste in una prova orale in cui sono discussi gli argomenti presentati nelle lezioni. Oltre all’apprendimento delle nozioni fondamentali esposte nel corso, vengono valutate anche le capacità e attitudini dello studente ad adattare i fondamenti teorici della chemiometria a particolari condizioni operative e pratiche; viene inoltre valutata la capacità espositiva e adeguatezza del linguaggio dello studente.

Per l'ammissione alla prova orale viene erogato un test a risposte multiple in aula informatica; ogni prova comprende 30 domande sugli argomenti esposti nelle lezioni del corso; gli studenti che ottengono esito positivo (almeno 60% risposte corrette) possono sostenere la prova orale. Per l'ammissione all'esame è obbligatorio aver frequentato le esperienze di laboratorio.

La prova orale è generalmente consecutiva al test al computer ma, su richiesta dello studente, può anche essere sostenuta in qualsiasi appello successivo alla prova scritta. Nel caso di non superamento della prova orale, non è previsto il salto d'appello e non è richiesta una nuova prova scritta.

## Orario di ricevimento

Previo appuntamento tramite e-mail, i docenti sono sempre disponibili a ricevere gli studenti nei loro uffici.