

Curso de Posgrado

Introducción a la exploración y análisis de datos multivariados

Contenidos

Bloque 1: Introducción al análisis multivariado de datos. Matriz de datos. Dimensión y estructura.

Aprendizaje estadístico supervisado y no supervisado. Regresión y clasificación. Modelos: selección y validación.

Bloque 2: Manejo de softwares de tratamiento de datos (Matlab –interfaces gráficas: MVC1, Regression toolbox, PCA toolbox y Classification toolbox-; SIMCA; Infostat, Metaboanalyst). Tipos de variables.

Creación de arreglos: vectores y matrices Ejecución de interfaces gráficas de usuario (GUI) en entorno Matlab.

Bloque 3: Datos químicos experimentales. Caracterización de datos de tipo multivariado. Preprocesado de datos. Manipulación de archivos de datos, carga e importación. Herramientas gráficas.

Bloque 4: Análisis exploratorio multivariado. Descomposición matricial y reducción de la dimensión.

Análisis de componentes principales (PCA). Concepto de variable latente. Variancia explicada. Matrices de loadings y scores. Interpretación gráfica. PCA como insumo para otras técnicas: diagnóstico de valores atípicos, regresión, agrupamiento [clustering].

Bloque 5: Modelos de regresión multivariada para datos químicos directos e inversos: regresión lineal múltiple (RLM), regresión en componentes principales (PCR) y regresión en cuadrados mínimos parciales (PLS). Estrategias de calibración. Selección de modelos. Errores. Validación cruzada.

Bloque 6: Modelos de clasificación multivariada. Modelado de clases: modelado suave independiente por analogía de clases (SIMCA). Análisis discriminante: lineal (LDA), cuadrático (QDA), por cuadrados mínimos parciales (PLS-DA). Evaluación de los modelos: validación cruzada, índices de clasificación, curvas ROC.

Bloque 7: Introducción a los modelos de machine learning. Fundamentos básicos de redes neuronales artificiales (ANN) y máquinas de soporte vectorial (SVM) para clasificación.

Bloque 8: Fusión de datos. Multiset. multi-block. Arreglos de matrices [matrix augmentation, matrix couple]. Niveles [bajo, medio y alto] y estrategias de fusión.

Metodología y evaluación

El dictado se desarrollará mediante presentaciones teóricas que proporcionarán una base sólida sobre los conceptos fundamentales del curso.

Por otra parte, se realizarán sesiones prácticas que permitirán la aplicación de los conceptos teóricos a problemas reales, facilitando el aprendizaje activo y la discusión en grupo. Estas sesiones se realizarán

principalmente mediante la práctica con herramientas y softwares relevantes para el campo de estudio, proporcionando habilidades técnicas aplicables en la práctica profesional. Adicionalmente, se implementará el uso de entornos virtuales para la entrega de material, foros de discusión y colaboración en línea.

Como método de evaluación se buscará evaluar el dominio de los conceptos como así también el manejo de software y herramientas específicas cubiertos durante el curso. Para ello, se entregará un trabajo final con casos prácticos a resolver aplicando los conceptos desarrollados, contando con 10 días posteriores a la finalización el curso para su resolución y entrega. La evaluación será aprobada con una calificación mínima de 6 [seis] y máxima de 10 [diez].

Carga horaria y cronograma

El curso tendrá una carga horaria de 40 horas [h] distribuidas en clases teóricas (T) y prácticas (P) y el trabajo evaluativo final como lo indica el siguiente cronograma:

	Día 1	Día 2	Día 3	Día 4	Día 5
Bloque 1	1.5 h - T				
Bloque 2	1.5 h - P				
Bloque 3	3 h - T-P				
Bloque 4		4 h - T-P			
Bloque 5		2 h - T	4 h - P		
Bloque 6			2 h - T-P	6 h - T-P	
Bloque 7					2 h T-P
Bloque 8					4 h T-P
Trabajo final evaluativo (10 h)					

Docentes responsables

Silvana Mariela Azcárate [Facultad de Ciencias Exactas y Naturales, Incitap/Conicet] es Licenciada [UNLPam] y Doctora en Química [UNSL]. Se especializa en análisis de muestras agroalimentarias, aplicación y estudios de algoritmos de clasificación y calibración, y análisis de datos. Es Miembro de la Asociación Argentina de Química Analítica [AAQA] desde el año 2013.

Daniela Alejandra Ortiz es Doctora en Agronomía [Facultad de Agronomía, Universidad Nacional del Sur Bahía Blanca. Licenciada en Química con Orientación en Alimentos. Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de La Pampa [UNLPam]. Participa de proyectos relacionados a la evaluación de calidad de materias primas y el desarrollo, validación y aplicación nuevas tecnologías para el análisis de diferentes forrajes y alimentos para consumo humano con enfoque al uso de tecnología NIRS y aplicación de herramientas quimiométricas.