

Las Malvinas
son argentinas



LICENCIATURA EN CIENCIA DE DATOS

Facultad de Ingeniería Química ¹

Facultad de Ingeniería y Ciencias Hídricas

Facultad de Ciencias Económicas

¹Sede sugerida

Universidad Nacional del Litoral



Introducción

Los avances tecnológicos e informáticos de nuestros tiempos, el acceso a grandes volúmenes de datos y el crecimiento vertiginoso de la capacidad de cálculo requieren de profesionales que dominen herramientas y técnicas que permitan comprender, sistematizar y analizar datos. Estos nuevos desafíos interpelan a las más variadas áreas del conocimiento en el proceso y demandan figuras con una sólida formación en matemática y estadística, con conocimientos específicos para abordar el manejo de datos, y con una gran capacidad de integración de distintos saberes para llevar a cabo aplicaciones concretas. La propuesta aquí presentada pretende cubrir esa demanda, proveyendo sólida formación y experiencia en Ciencia de Datos.

El énfasis será puesto en que cada estudiante aprenda a utilizar métodos estadísticos para analizar grandes conjuntos de datos, programar los modelos predictivos, y comprender y desarrollar aplicaciones estratégicas para la toma de decisiones.

Para ello, en cada una de las asignaturas del ciclo básico los temas a abordar serán introducidos y desarrollados utilizando datos de diferente naturaleza y áreas, incluyendo, pero sin ser exhaustivos, datos económicos, hídricos, biomédicos, químicos, entre otros.

Por otro lado el objetivo de los cursos optativos es que cada estudiante domine técnicas y algoritmos eficientes y potentes para extraer información válida de datos en las áreas que quiera lograr su experticia.



Antecedentes y fundamentación

La Ciencia de Datos se encarga de estudiar, analizar y extraer información de diferentes tipos de datos. Se trata de un área interdisciplinaria, ya que combina la informática con la estadística y las matemáticas para procesar datos.

La Universidad Nacional del Litoral cuenta con carreras de matemática, informática y física de muy alto nivel pero no con carreras que vinculen las tres disciplinas y con una visión que incluya la preparación para el ámbito empresarial e industrial. En la Universidad de Buenos Aires acaba de crearse la carrera de Licenciatura en Ciencia de Datos con características similares a la presentada en este documento.

Este escenario demuestra la necesidad de crear una nueva carrera en la Universidad Nacional del Litoral que provea sólidos conocimientos integrados en matemática, computación, estadística y una fuerte formación científica-tecnológica y profesional que permita una rápida en el sector público, productivo y de servicios, así como también interactuar con grupos de investigación interdisciplinarios o continuar una formación de posgrado.

En los últimos años, algunos países de Europa midieron el impacto de la formación en matemática. Los resultados son coincidentes y concluyen que la matemática es un "conocimiento estratégico que genera alto valor agregado en virtualmente todos los sectores de la economía". En Argentina no se cuenta con informes de este tipo y el crecimiento que se ha dado en carreras vinculadas a la matemática en Europa no ha sido observado aún en el ámbito local, posiblemente debido a la falta de ofertas académicas.

Por este motivo se pretende en el primer ciclo una etapa de formación con una fuerte base de matemática, estadística y computación.

Por otro lado, el área de ciencia de datos está en permanente desarrollo, apareciendo continuamente nuevos métodos y áreas de aplicación. Por ello se propone un segundo ciclo flexible, que contempla varias especializaciones posibles, de acuerdo al interés y preferencia individuales de cada estudiante. Teniendo como horizonte la ampliación de áreas de especialización se realizarán esfuerzos para que cada estudiante pueda cursar asignaturas optativas en centros o casas de estudios de nuestra propia universidad y en otros con los que se alcancen o ya presenten convenios de cooperación para tal fin.



Objetivos de la Carrera

La carrera Licenciatura en Ciencia de Datos tiene como objetivos principales:

1. Proveer una amplia y sólida formación en matemática, estadística y computación, generando capacidades para afrontar y resolver desafíos provenientes tanto del sector productivo como del ámbito académico, en distintas áreas que requieran del modelado y manejo de datos.
2. Cubrir la creciente demanda por parte de instituciones públicas y privadas de recursos humanos con fuerte formación cuantitativa, solidez en programación, pensamiento crítico, actitud científica, capacidad de abstracción y destreza lógica.

La formación adquirida en la carrera otorgará amplios conocimientos en matemática, estadística y ciencias de la computación. Esa formación disciplinar junto con una gran capacidad de abstracción, razonamiento lógico y pensamiento crítico permitirán una inserción exitosa en el ámbito profesional, en desarrollos e investigaciones que requieran de herramientas rigurosas para el análisis e interpretación de datos. También brindará las condiciones necesarias para continuar estudios de posgrado que requieran de una sólida formación teórica y metodológica en matemática, estadística y computación. Las principales capacidades que tendrán los graduados serán:

1. Riguroso manejo de la matemática detrás de los métodos en Ciencia de Datos;
2. Gran capacidad para el análisis y el desarrollo de algoritmos para implementar dichos métodos;
3. Capacidad para ejecutar y adaptar los métodos y algoritmos en aplicaciones concretas a diversas áreas.



Descripción de la carrera

Denominación de la carrera

Licenciatura en Ciencia de Datos

Título a otorgar

Licenciado o Licenciada en Ciencia de Datos

Duración

Cinco (5) años.

Requisitos de ingreso

Cumplimentar con los requisitos de ingreso que estipule la Universidad Nacional del Litoral.

Sede

Facultad de Ingeniería Química

Facultades intervinientes

Facultad de Ingeniería Química

Facultad de Ingeniería y Ciencias Hídricas

Facultad de Ciencias Económicas



Perfil del título y alcances

Quien egrese de la Licenciatura en Ciencia de Datos contará con una sólida formación interdisciplinaria en matemáticas y computación, que le brindará habilidades avanzadas en modelado matemático, programación, análisis estadístico, aprendizaje automático, minería, limpieza y manipulación de datos, visualización y generación de informes, análisis de riesgo y visión de negocio, herramientas en la nube, estructuras y almacenamiento de datos, entre otras.

Los rasgos por los que se distingue un/a Licenciada/o en Ciencia de Datos, son la capacidad de abstracción y modelado que le permite no sólo abordar un problema de gestión y explotación de datos, sino también contar con las aptitudes necesarias para diseñar nuevos algoritmos o experimentos que dirijan la explotación hacia nuevos resultados que puedan ser de interés para la compañía/institución en la que se desempeñe.

Su desempeño como profesional podrá llevarse a cabo en empresas privadas o estatales, o a través de consultorías, aplicando técnicas matemáticas, estadísticas y computacionales en la solución de problemas vinculados a áreas tan diversas como la economía, negocios, administración, biología, genómica, química, agronomía, ecología, medicina y demografía, entre otras.

Su campo ocupacional también incluye la investigación y docencia en temas afines en Universidades, Organismos de Investigación públicos y/o privados, en áreas interdisciplinarias que vinculen la matemática y la computación.

A continuación, se detallan los Alcances del Título:

- Diseñar, implementar, ejecutar y validar modelos, aplicando metodologías y principios matemáticos, computacionales y estadísticos vinculados al tratamiento de datos.
- Brindar asesoramiento para la aplicación de teorías, metodologías y recursos de la Ciencia de Datos en actividades productivas y de servicios.
- Realizar tareas de consultoría, auditoría, inspecciones y/o pericias legales que requieran especialistas en Ciencias Matemáticas, de la Computación y de Datos en cuerpos ejecutivos, legislativos y judiciales, u otras instituciones oficiales o privadas.
- Participar en proyectos de investigación científica, técnica y tecnológica interdisciplinarios en las distintas áreas de su competencia.
- Participar en el diseño, desarrollo y evaluación de documentos y materiales sobre temas de las distintas áreas de las Ciencias Matemáticas, de la Computación y de Datos, organizando y administrando programas de desarrollo disciplinar para la formación de recursos humanos en los distintos niveles educativos, así como para el público en general.



Plan de estudios

Características del plan de estudios

El paradigma bajo el cual se formula este Plan de Estudios puede resumirse en los siguientes principios:

- Planificación de una carrera de grado de cinco años de duración con salida laboral en el ámbito académico, industrial y empresarial.
- Organización de los primeros años de cursado de manera que se contemple la problemática actual de deserción de estudiantes de carreras de grado. Esto se ve reflejado en la carga horaria que va decreciendo con los años ya que se pretende que el cursado sea más personalizado en un comienzo y otorgarle herramientas que, a medida que avance en el cursado, le permitan trabajar en forma más independiente.
- Materias en común con estudiantes de otras carreras de la Facultad de Ingeniería Química y la Facultad de Ingeniería y Ciencias Hídricas, con el objetivo de favorecer la formación interdisciplinaria.
- Énfasis en el uso de software libre.
- Tratamiento de los contenidos fundamentales desde diversos puntos de vista: conceptual, histórico, numérico y/o gráfico.
- Organización de la carrera en dos ciclos: inicial y superior. Contempla el cursado de asignaturas obligatorias, optativas y la realización de un Trabajo Final.

Todas las asignaturas serán cuatrimestrales, excepto la asignatura Trabajo Final, que será anual. El plan de estudios constará de dos ciclos con un total de 2880 horas de cursado.



Ciclo básico

Algunos cursos del ciclo básico serán comunes con cursos de la Licenciatura en Matemática Aplicada de la FIQ y otros con cursos de Ingeniería Informática y otras carreras relacionadas de la FICH. El ciclo básico suma un total de 1920 horas de cursado.

Las asignaturas que formarán parte de este ciclo son:

- Álgebra
- Matemática Básica
- Ciencia de datos I
- Matemática Discreta
- Cálculo I
- Ciencia de datos II
- Algoritmos y estructuras de datos
- Álgebra lineal
- Ciencia de datos III
- Análisis Matricial
- Cálculo II
- Bases de datos
- Probabilidad
- Métodos numéricos I
- Optimización y aprendizaje automático I
- Estadística
- Métodos numéricos II
- Optimización y aprendizaje automático II



Ciclo superior

El ciclo superior constará de cursos específicos elegidos entre diferentes áreas y un Trabajo Final de carrera.

Los cursos del ciclo superior no son fijos y dependerán de la oferta académica que decidirá cada cuatrimestre el Comité Académico teniendo en cuenta las disponibilidades de recursos humanos. Las materias optativas deben sumar un total de al menos 720 horas de cursado. Además cada una de ellas deberán tener un mínimo de 60 horas y un máximo de 120 horas. Además teniendo que cumplimentar un mínimo en las siguientes tres áreas:

Área Fundamentos de la estadística (al menos 90 horas)

Área Métodos para el análisis y procesamiento de datos (al menos 180 horas)

Área Programación (al menos 90 horas)

A las que se pueden sumar optativas del **Área Aplicaciones Específicas de ciencias de datos**.

Se presenta en [Sección Optativas](#), algunos ejemplos orientativos de cursos en cada una de estas áreas y del área aplicaciones específicas de ciencias de datos.

El trabajo final tiene una carga horaria de 240 horas y es anual.

Requisito de idioma extranjero

Se deberá acreditar conocimiento de Idioma Extranjero: inglés, conforme a la normativa vigente en la Universidad Nacional del Litoral.

Organización del cursado

A continuación, se presenta un cuadro que muestra un modelo de cursado para la Licenciatura en Ciencia de Datos. La distribución de los cursos en cuatrimestres sigue un posible esquema temporal. En el [Anexo](#) se expone un sistema de correlatividades, que permite otros reordenamientos, siempre respondiendo a criterios que promueven la articulación de saberes y contenidos entre las diversas áreas de las disciplinas matemática, estadística y computación en el proceso de formación, a la vez que atienden la complejidad de los niveles y los requerimientos de conocimientos para el avance en el cursado y aprobación de las asignaturas.



Distribución de asignaturas por cuatrimestre

Año	Cuat.	Código	Asignatura	Horas	
				Semanales	Totales
Ciclo Inicial					
1ro.	I	1	Álgebra (FIQ)	8	120
		2	Matemática Básica (FIQ)	8	120
		3	Ciencia de datos I (FCE-FICH-FIQ) *	6	90
	II	4	Matemática Discreta (FIQ)	10	150
		5	Cálculo I (FIQ)	8	120
		6	Ciencia de datos II (FCE-FICH-FIQ) *	6	90
2do.	III	7	Algoritmos y estructuras de datos (FICH)	6	90
		8	Álgebra Lineal (FIQ)	8	120
		9	Ciencia de datos III (FCE-FICH-FIQ) *	6	90
	IV	10	Análisis matricial (FIQ)	6	90
		11	Cálculo II (FIQ)	8	120
		12	Bases de datos (FICH)	4	60
3ro.	V	13	Probabilidad (FIQ)	6	90
		14	Métodos numéricos I (FIQ)	6	90
		15	Optimización y aprendizaje automático I (FIQ) *	6	90
	VI	16	Estadística (FIQ)	8	120
		17	Métodos numéricos II (FIQ)	6	90
		18	Optimización y aprendizaje automático II (FIQ) *	6	90
Ciclo Superior					
4to.	VII		Optativa	6	90
			Optativa	6	90
	VIII		Optativa	6	90
			Optativa	6	90
5to.	IX		Trabajo Final (anual) *	8	240
			Optativa	6	90
			Optativa	6	90
	X		Optativa	6	90
			Optativa	6	90
			Optativa	6	90
Carga horaria total de la carrera				2880	

* A implementar



Optativas

Presentamos a continuación ejemplos de materias optativas a dictarse en las diferentes áreas.

Área Fundamentos de la estadística (al menos 90 horas)

- Análisis estadístico de datos funcionales (FIQ)
- Análisis multivariado I (FIQ)
- Análisis multivariado II (FIQ)
- Estadística en alta dimensión (FIQ)
- Estimación no paramétrica (FIQ)
- Modelos lineales generalizados (FIQ)
- Teoría asintótica (FIQ)

Área Métodos para el análisis y procesamiento de datos (al menos 180 horas)

- Diseño y análisis de experimentos (FIQ)
- Inteligencia Artificial (FICH)
- Metodología de superficies de respuesta (FIQ)
- Métodos estadísticos no paramétricos (FIQ)
- Modelos lineales (FIQ)
- Señales y sistemas (FICH).
- Procesamiento digital de imágenes (FICH)
- Redes neuronales I (FICH)
- Redes neuronales II (FICH)
- Regresión y técnicas multivariadas (FIQ)

Área Programación (al menos 90 horas)

- Computación gráfica (FICH)
- Ingeniería de software (FICH)
- Desarrollo ágil de software (FICH)



-
- Administración de Proyectos de Software (FICH)
 - Programación concurrente y paralela (FICH)
 - Metaheurísticas (FICH)
 - Sistemas operativos (FICH)
 - Tecnologías de programación (FICH)

Área Aplicaciones Específicas de ciencias de datos

- Ciencia de datos en Economía y Negocios (FCE)
- Análisis de datos socioeconómicos (FCE)
- Ciencia de Datos aplicado a la Industria (FIQ)
- Interfaces humano-computadora (FICH)
- Métodos Econométricos I: Fundamentos (FCE)
- Métodos Econométricos II: Series de Tiempo (FCE)
- Problemas Inversos con aplicaciones al procesamiento de imágenes digitales (FIQ)
- Procesamiento del lenguaje natural (FICH)
- Visión por computadora (FICH)



Desarrollo por asignatura: obligatorias

Álgebra

Objetivos

Lograr un conocimiento acabado sobre los conjuntos numéricos fundamentales, como así también adquirir técnicas de escritura y formalización de demostraciones matemáticas.

Contenidos mínimos

Conjuntos, funciones. Números reales, propiedades de cuerpo ordenado. Números naturales, inducción. Números enteros, divisibilidad. Números racionales. Propiedad de completitud. Números complejos. Polinomios, divisibilidad, raíces. Introducción a la escritura de textos matemáticos y al sistema LaTeX.

Carga horaria semanal

8 horas.

Matemática Básica

Objetivos

Introducir las herramientas básicas necesarias que permitan comprender y manipular adecuadamente los conceptos y las técnicas del Cálculo y del Álgebra Lineal.

Contenidos mínimos

Valor absoluto, ecuaciones y desigualdades. Trigonometría. Funciones y gráficas. Funciones trigonométricas, exponenciales y logarítmicas. Cónicas. Sistemas de ecuaciones. Vectores y matrices. Determinantes.

Carga horaria semanal

8 horas.



Ciencia de datos I

Objetivos

Introducción a la algorítmica y estadística de la Ciencia de Datos. Es una mirada integral a los problemas con datos, donde se dará la misma importancia a los fundamentos de programación, manejo de datos, modelos y estadística asociados a los mismos. Se introducirán los temas más básicos de programación, visualización de datos y análisis estadístico serán incluidos simultáneamente, es decir, se avanzará en cada clase con contenidos de las tres áreas. Los contenidos a abordar con esta mirada integral serán introducidos y desarrollados desde el planteo de un problema que tenga sentido práctico y esté al alcance para estudiantes del ciclo inicial, utilizando datos de diferente naturaleza y áreas, incluyendo por ejemplo, datos económicos, hídricos, biomédicos, químicos, etc.

Contenidos mínimos

Algoritmos y programas. Lenguajes, compiladores e intérpretes. Pseudocódigo y diagramas de flujo. Identificadores, constantes, variables y expresiones. Tipos de datos. Estructuras de control condicionales y repetitivas, anidación. Programación estructurada. Funciones. Variables cuantitativas y cualitativas. Variables predictoras, variables respuesta. Relaciones entre variables. Visualización. Análisis exploratorios basados en las preguntas de interés. Población y muestras.

Carga horaria semanal

6 horas.

Ciencia de datos II

Objetivos

Desarrollo de la algorítmica y estadística de la Ciencia de Datos. Es una mirada integral a los problemas con datos, donde se dará la misma importancia a los fundamentos de programación, manejo de datos, modelos y estadística asociados a los mismos. Se continuará con el desarrollo de los temas de programación, visualización de datos y análisis estadístico serán incluidos simultáneamente, es decir, se avanzará en cada clase con contenidos de las tres áreas. Los contenidos a abordar con esta mirada integral serán introducidos y desarrollados desde el planteo de un problema que tenga sentido práctico y esté al alcance para estudiantes del ciclo inicial, utilizando datos de diferente naturaleza y áreas, incluyendo por ejemplo, datos económicos, hídricos, biomédicos, químicos, etc.



Contenidos mínimos

Arreglos y matrices. Listas. Recorridos y búsqueda. Algoritmos recursivos. Archivos. Adquisición de datos. Limpieza de datos. Visualización de relaciones entre variables. Introducción a modelos. Correlación. Comparación de medias. Test de hipótesis. Regresión. Bootstrap. Remuestreo. Test de permutación. Simulación de datos.

Carga horaria semanal

6 horas.

Ciencia de datos III

Objetivos

Profundización de la algorítmica y estadística de la Ciencia de Datos. Es una mirada integral a los problemas con datos, donde se dará la misma importancia a los fundamentos de programación, manejo de datos, modelos y estadística asociados a los mismos. Se profundizarán los temas de programación, visualización de datos y análisis estadístico serán incluidos simultáneamente, es decir, se avanzará en cada clase con contenidos de las tres áreas. Los contenidos a abordar con esta mirada integral serán introducidos y desarrollados desde el planteo de un problema que tenga sentido práctico y esté al alcance para estudiantes del ciclo inicial, utilizando datos de diferente naturaleza y áreas, incluyendo por ejemplo, datos económicos, hídricos, biomédicos, químicos, etc.

Contenidos mínimos

Programación orientada a objetos: clases, instancias, mensajes, herencia y agregación, encapsulamiento, polimorfismo. Documentación, mantenimiento, desarrollo en equipos y sistemas de control de versiones. Reusabilidad, reproducibilidad, replicabilidad y accesibilidad. Estimación no paramétrica de densidad. Introducción a regresión lineal múltiple. Introducción a modelos no lineales. Modelado de relación entre variables cualitativas. Introducción a los modelos lineales generalizados. Residuos. Diagnóstico. Medidas de errores.

Carga horaria semanal

6 horas.



Bases de datos

Objetivos

Lograr conocer el manejo de bases de datos relacionales y el lenguaje de consultas estándar, la administración de motores de bases de datos y la transformación de modelos conceptuales entidad-relación.

Contenidos mínimos

Introducción a las bases de datos. Transformación de modelos conceptuales a modelos físicos de datos. Normalización de bases de datos. Transacciones. Bases de datos relacionales y no relacionales.

Carga horaria semanal

4 horas.

Matemática Discreta

Objetivos

Aprender los fundamentos de la teoría de números y combinatoria, como así también desarrollar destrezas para su utilización en la resolución de problemas teóricos y sus aplicaciones actuales.

Contenidos mínimos

Repaso de conjuntos y definiciones recursivas. Relaciones: órdenes, equivalencias y funciones. Funciones inyectivas, sobre y biyectivas. Principio de las cajas. Conteo. Principios enumerativos. Selecciones no ordenadas con repetición. Teorema del binomio. Principio de la Criba. Particiones, clasificaciones y distribuciones. Aritmética modular. Congruencias, aritmética en \mathbb{Z}_m . Ecuaciones diofánticas. Grafos y árboles. Algoritmos. Búsqueda en árboles.

Carga horaria semanal

10 horas.



Cálculo I

Objetivos

Adquirir los conceptos esenciales del cálculo diferencial e integral de funciones de una variable y la habilidad para usar su lenguaje y sus técnicas en problemas analíticos, geométricos y físicos.

Contenidos mínimos

Sucesiones. Series. Límite y continuidad de funciones. Derivadas. Valores extremos de funciones. Integrales. Cálculo de primitivas. Integral definida. Integrales impropias. Polinomios de Taylor, series de potencias. Implementaciones en software matemáticos para la visualización y resolución de problemas.

Carga horaria semanal

8 horas.

Álgebra Lineal

Objetivos

Familiarizarse con los aspectos geométricos y vectoriales de los espacios euclídeos como asimismo con las aplicaciones lineales definidas entre ellos y comprender adecuadamente estas nociones en el contexto de espacios vectoriales más generales.

Contenidos mínimos

Vectores, rectas y planos. Producto escalar, proyecciones y producto vectorial. Espacios vectoriales. Transformaciones lineales. Autovalores y autovectores. Espacios euclídeos, espacios hermíticos, formas bilineales y cuadráticas.

Carga horaria semanal

8 horas.

Análisis matricial

Objetivos

Aprender resultados clásicos y recientes del análisis matricial que resultan centrales en diversas aplicaciones (Ecuaciones Diferenciales, Probabilidad, Estadística, Optimización)



y que contienen resultados que escapan los alcances de un curso de Álgebra Lineal elemental.

Contenidos mínimos

Repaso de autovalores/autovectores. Matrices normales y unitarias. Teorema de triangulación de Schur y aplicaciones. Factorización QR. Factorizaciones triangulares (LU, Cholesky). Matrices hermíticas, cociente de Rayleigh. Diagonalización simultánea de matrices simétricas y hermíticas. Normas de vectores y normas inducidas de matrices. Propiedades algebraicas y analíticas de las normas. Radio espectral. Forma polar y descomposición en valores singulares. Desigualdades para matrices definidas positivas.

Carga horaria semanal

6 horas.

Cálculo II

Objetivos

Adquirir los conceptos esenciales del análisis de funciones de varias variables con valores reales y vectoriales, el manejo de sus técnicas y la aplicación de sus resultados a problemas físicos.

Contenidos mínimos

Funciones vectoriales. Longitud de arco. Curvatura, torsión, plano osculador. Derivadas. Funciones de varias variables. Derivadas parciales. Regla de la cadena. Derivadas direccionales. Extremos y extremos condicionados. Integración múltiple. Aplicaciones. Integrales de línea. Integrales de superficie. Teoremas de Green, Gauss y de Stokes.

Carga horaria semanal

8 horas.

Algoritmos y estructuras de datos

Objetivos

Adquirir un conocimiento exhaustivo de las principales estructuras de datos y aprender a implementarlas en forma eficiente; analizar diferentes algoritmos de acceso y manejo a tales estructuras de datos, utilizando un formalismo matemático para estimar la eficiencia de los algoritmos.



Contenidos mínimos

Listas vinculadas simples y dobles. Pilas, filas. Estructuras de Datos no lineales con árboles. Recursión. Algoritmos sobre grafos. Elementos básicos de complejidad de algoritmos. Complejidad temporal y espacial. Técnicas de diseño de algoritmos: divide y conquista, Programación dinámica, greedy, backtracking. Problemas NP-Hard.

Carga horaria semanal

6 horas.

Probabilidad

Objetivos

Familiarizarse con el andamiaje formal riguroso de la teoría matemática de probabilidades.

Contenidos mínimos

Espacios de probabilidad. Independencia. Probabilidad Condicional. Teorema de Bayes. Variables Aleatorias. Distribuciones discretas, continuas y singulares. Función de distribución acumulativa. Distribuciones de funciones de variables y vectores aleatorios. Distribuciones más usuales. Esperanza de una variable aleatoria y sus propiedades. Momentos. Varianza y sus propiedades. Covarianza y correlación. Distribución y esperanza condicionales. Sucesiones de variables aleatorias. Convergencia casi segura y convergencia en probabilidad. La Ley de los grandes números. El lema de Borel-Cantelli. Convergencia en distribución. Teorema central del límite.

Carga horaria semanal

6 horas.

Métodos Numéricos I

Objetivos

Se brinda una introducción al cálculo numérico y en especial se estudian métodos para resolver ecuaciones y sistemas de ecuaciones lineales y no lineales algebraicas. Con esto, se pretende que cada estudiante adquiera habilidad en la resolución numérica de problemas, a través de una introducción moderna de los conceptos y de numerosos ejemplos, implementando los métodos en computadoras. Además, se busca que inicien el desarrollo de su capacidad de análisis y comprendan los resultados experimentales obtenidos.



Contenidos mínimos

Representación en punto flotante, redondeo, errores. Error absoluto y error relativo. Resolución de ecuaciones no-lineales. Método de Bisección. Método de punto fijo. Método de la secante. Método de Newton. Análisis del error y orden de convergencia. Resolución de sistemas de ecuaciones no-lineales. Métodos de punto fijo y de Newton. Análisis del error y orden de convergencia. Interpolación polinomial. Estimación del error. Aproximación por cuadrados mínimos. Interpolación polinomial a trozos. Integración numérica. Fundamentos, métodos y estimación del error. Cálculo aproximado de integrales.

Carga horaria semanal

6 horas.

Métodos Numéricos II

Objetivos

Se brindan métodos avanzados de cálculo numérico. Profundización de los temas desarrollados en Métodos Numéricos I. Con esto, se pretende que cada estudiante profundice su habilidad en la resolución numérica de problemas, a través de una introducción moderna de los conceptos y de numerosos ejemplos, implementando los métodos en computadoras. Además, se busca que desarrollen una capacidad avanzada de análisis y comprendan con alto nivel de detalle los resultados experimentales obtenidos para saber en qué casos conviene utilizar un método u otro, conocer las ventajas y desventajas de cada uno y distinguir en qué casos algún método puede conducir a una solución errónea.

Contenidos mínimos

Métodos iterativos para sistemas de ecuaciones lineales. Jacobi, Gauss-Seidel, SOR, Richardson. Métodos iterativos para matrices SDP. Descenso más pronunciado, gradiente conjugado. Gradiente conjugado para problemas no simétricos. GMRes. Precondicionadores genéricos: diagonal, Gauss-Seidel, factorización incompleta. Gradiente conjugado precondicionado. Métodos numéricos para problemas de autovalores. Repaso. Método de las potencias. Método de las potencias inversas y algoritmo QR.

Carga horaria semanal

6 horas.



Estadística

Objetivos

Estudiar métodos estadísticos para el análisis de datos. Interpretar resultados obtenidos en la aplicación de un modelo estadístico determinado. Estimar distribuciones y testear hipótesis acerca de parámetros. Distinguir relaciones estadísticas que afectan a variables n-dimensionales.

Contenidos mínimos

La normal multivariada. El teorema central del límite multivariado para muestras independientes y para muestras no independientes. Estimación puntual y por intervalos de confianza para modelos lineales. Estudio asintótico de estimadores puntuales para modelos lineales. Modelos lineales generalizados.

Carga horaria semanal

8 horas.

Optimización y aprendizaje automático I

Objetivos

optimización utilizados en aprendizaje automático. Se introducirán los métodos de aprendizaje automático que utilizan algoritmos como herramienta de optimización. Luego se utilizarán dichos métodos para dar solución concreta a problemas reales y la construcción de modelos predictivos.

Contenidos mínimos

Métodos de optimización: optimización de funciones convexas, funciones convexas con penalización (lasso, ridge, elastic net). Gradiente descendiente y estocástico. Retropropagación. Aprendizaje supervisado: Regresión lineal. Regresión logística. Análisis discriminante lineal y cuadrático. Vecinos más cercanos. Árboles de regresión y de clasificación. Ensamblados. Métodos de selección y validación de modelos. Medidas de desempeño. Curvas ROC.

Carga horaria semanal

6 horas.



Optimización y aprendizaje automático II

Objetivos

Profundización de los fundamentos matemáticos y la implementación de métodos de optimización utilizados en aprendizaje automático. Por otro lado, se desarrollarán nuevos métodos de aprendizaje automático que utilizan dichos algoritmos como herramienta de optimización. Estos nuevos métodos se emplearán para dar solución concreta a problemas reales y la construcción de modelos predictivos.

Contenidos mínimos

Métodos de optimización: Maximización de funciones de verosimilitud. Problemas duales para maximizar. Algoritmo de maximización de la esperanza. Fundamentos de derivación automática para el aprendizaje profundo. Máquinas de vectores soporte. Aprendizaje no supervisado: Reducción de dimensiones. Análisis de componentes principales. Métodos de clustering. Biclustering. Aprendizaje semi-supervisado.

Carga horaria semanal

6 horas.

Trabajo Final

Objetivos

El trabajo final de la carrera de Licenciatura en Ciencia de Datos tiene por finalidad actuar como una primera experiencia entre el o la estudiante y su fuente de trabajo como Licenciada o Licenciado, favoreciendo un espacio de vinculación con empresas, institutos y otras facultades y universidades. Se pretende que funcione como un espacio en el que el o la estudiante pueda utilizar los conocimientos adquiridos a lo largo de su carrera académica para enfrentarse a problemas aplicados o teóricos, a fin de que pueda afianzarlos en un marco común. El software desarrollado y el procesamiento de datos debe estar disponible en una plataforma accesible y debe desarrollarse bajo los principios de reusabilidad, reproducibilidad, replicabilidad y accesibilidad.

Contenidos mínimos

Metodología de la investigación en Ciencia de Datos: Tipos de trabajos de investigación: investigación teórica, aplicada y aplicaciones. Direccionalidad teórica y consecuencias metodológicas. Etapas en un trabajo de investigación. Estrategias en la búsqueda y selección de la bibliografía.



Introducción general a la ética. Ética de datos. Transparencia, responsabilidad, privacidad y vulneración de derechos. Equidad algorítmica: criterios, métricas y algoritmos. Regulaciones y marcos de referencia normativos en distintos países.

Elaboración del proyecto: El proceso de investigación. Problematización inicial del tema. Herramientas para la elaboración de un plan de investigación: selección de objetivos, estado actual del arte, planteo de hipótesis y problemas, metodologías para abordar el trabajo propuesto.

Elaboración del trabajo final: Desarrollo de pautas formales para la redacción de trabajos escritos de investigación en Ciencia de Datos. Organización de los contenidos y estrategias de redacción del trabajo final. La planificación de los tiempos y las actividades. Exposición de los avances en la elaboración del proyecto de tesis.

Desarrollo del software de análisis y procesamiento de datos.

Redacción y composición de ponencias orales. Estrategias para la presentación y defensa oral de trabajos. Exposición y defensa del trabajo final.

Carga horaria semanal

8 horas.



Desarrollo por asignaturas: optativas

Análisis de datos socioeconómicos

Características de los datos económicos y sociales. Análisis de Datos de Encuesta. Visualización. Índices y reducción de dimensiones. Análisis factorial y de correspondencias múltiples. Aplicaciones socioeconómicas de Clustering. Análisis de Tablas de Contingencia. Modelos para Datos Categóricos: Respuesta dicotómica y multinomial. Modelos de utilidad aleatoria y aplicaciones. Enfoque de variables latentes. Modelos con variables de conteo. Datos censurados y truncados. Análisis de Datos Textuales. Redes.

Análisis estadístico de datos funcionales

Descripción de distintos procedimientos para analizar datos funcionales. Definición de diversos métodos de estimación y descripción de dificultades técnicas originadas en el hecho de trabajar en espacios de Hilbert. Nociones de Probabilidad en Espacios de Banach y de Hilbert. Estadística de datos funcionales. Predicción para datos funcionales. Correlación canónica, Clasificación y Análisis Discriminante de datos funcionales. Componentes principales funcionales. Modelos Lineales Funcionales para respuestas escalares.

Análisis multivariado I

Distribución normal multivariada. Distribuciones condicionales. Distribución Wishart. Distribución de los estimadores de máxima verosimilitud. Distribución de Hotelling. Test de Hotelling. Aplicación al problema de dos muestras y al análisis de la varianza. Estadístico U de Rao. Aplicación del test de Hotelling. Problema de Fisher-Behrens. Análisis de perfiles. Métodos de reducción de dimensión. Componentes principales. Biplot. Coordenadas discriminantes. Clasificación y análisis discriminante. Clasificación no paramétrica.



Análisis multivariado II

Aspectos teóricos de componentes principales. Curvas principales. Biplot. Correlación canónica. Discriminación y clasificación. Cluster.

Ciencia de Datos aplicada a la Industria

Ingeniería matemática en los negocio y soporte en la toma de decisión en ambientes empresariales. Resolución de problemas específicos de distintas industrias utilizando herramientas de Optimización, programación matemática estocástica, aprendizaje estadístico, inteligencia de negocios y ciencia de datos, mediante el abordaje computacional, el análisis y la resolución.

Ciencia de datos en Economía y Negocios

Fundamentos del *Business Analytics* y de *Business Intelligence*. Data cleansing y análisis exploratorio. Análisis de datos con SQL. Utilización de métodos de machine learning y *deep learning* en economía y negocios. Modelos predictivos aplicado a ciencias económicas, empresariales y sociales. *Business Analytics* para la toma de decisiones. Elementos de visualización, presentación y reporte de resultados para potenciales usuarios.

Computación gráfica

Análisis de modelos y formas de representación gráfica de estructuras bidimensionales y tridimensionales. Conceptos y algoritmos de computación gráfica, dispositivos gráficos, geometría y algoritmos para objetos 2-D y 3-D, transformaciones, Iluminación y sombreado, Color, aspectos de renderizado de imágenes realistas.

Diseño y análisis de experimentos

Modelos lineales. Estimación y distribución de los estimadores. Test para los parámetros del modelo. Regresión. Inferencia y predicción. Experimentos con un factor. Análisis de la varianza. Comparaciones múltiples. Validación del modelo. Modelos de efecto fijo y efecto aleatorio. Diseños con bloques. Diseño de experimentos multifactorial. Diseños mixtos. Diseños anidados. Datos no balanceados. Diseño factorial 2k. Confusión en el diseño factorial 2k. Diseños fraccionales. Superficie de respuesta.



Estadística en alta dimensión

Modelos lineales en alta dimensión. Estimación de grandes matrices de covarianza. Reducción de dimensiones. Selección de modelos. Fenómenos de concentración y teoría no asintótica. Resultados básicos de la teoría de matrices aleatorias.

Estimación no paramétrica

Estimación no paramétrica de la Densidad: Estimación por Núcleos, Error Cuadrático Medio, Error Cuadrático Medio Integrado. Regresión No Paramétrica: Modelos no paramétricos, estimación por núcleos, polinomios locales, vecinos más cercanos, método de splines. Estimación de la Derivada. Selección del parámetro de suavizado. inferencia con regresión no paramétrica. Caso Multivariado. Validación cruzada, funciones de penalización. Método plug-in. Datos con outliers. M-smoothing. Introducción a los modelos semiparamétricos y aditivos.

Ingeniería de software

Ciclo de desarrollo de software. Metodologías ágiles. Lenguaje de Modelado UML. Diseño de software. Patrones de diseño. Arquitectura de sistemas. Patrones arquitectónicos. Estándares de calidad de software.

Inteligencia artificial

Agentes, búsqueda, inferencia lógica, representación del conocimiento, planificación. Sistemas expertos. Aprendizaje por refuerzo.

Interfaces humano-computadora

Interfaces basadas en el habla: reconocimiento automático del habla, diferentes modalidades, problemas y enfoques. Computación afectiva. Procesamiento cognitivo del habla y señales relacionadas. Interfaces cerebro-computadora: principales desafíos, paradigmas y métodos de registro.

Desarrollo ágil de software

Complejidad de los proyectos. La agilidad como nueva visión. Requerimientos Ágiles. Extreme Programming. Scrum. Kanban. Management Ágil.



Administración de proyectos de software

Concepto de Proceso, utilización y madurez. Gestión de Requerimientos. Evaluación de Proyectos. Planeamiento de Proyecto. Partición estructurada (WBS) y Estimación. Planeamiento de calendarios de Proyecto. Plan de Riesgos. Monitoreo y control. Gestión organizacional del Proyecto. Recursos humanos. Gestión tecnológica del Proyecto. Configuración y herramientas. Gestión de la Calidad.

Metaheurísticas

Concepto de metaheurística y espacio de soluciones. Metaheurísticas basadas en poblaciones, trayectorias y en adaptación social: fundamento, representación de la solución, operadores y evaluación de la calidad de las soluciones. Algoritmos meméticos. Adaptación de parámetros y operadores. Optimización multiobjetivo.

Metodología de superficie de respuesta

Introducción a la metodología de análisis de superficie de respuesta. Diseños factorial completo, central compuesto, Box-Behnken, Doehlert, optimal y de mezclas. Evaluación de los modelos con ajuste por cuadrados mínimos. Análisis de gráficas de superficie y de contorno. Respuestas múltiples. Función deseabilidad. Utilización del programa Design Expert y/o JMP Software.

Métodos econométricos I: Fundamentos

Fundamentos de la Econometría. Diferenciación entre econometría y estadística. Ejemplos reales motivadores. Revisión de álgebra matricial. El Modelo de Regresión Lineal Múltiple en el marco de los modelos económicos. Estimación mínimo cuadrática y propiedades de los estimadores. Omisión de Variables Explicativas e Inclusión de Irrelevantes. Tests de Restricciones Lineales. Problemas de Multilinealidad, Heteroscedasticidad y Autocorrelación. Mínimos Cuadrados Generalizados. Criterios de Evaluación de Modelos Econométricos. Estabilidad de los parámetros. Inclusión de covariables dummies. Aplicaciones de la Teoría Asintótica: el método 'delta'. Introducción a Máxima Verosimilitud. Endogeneidad y Variables Instrumentales (IV). Modelos con ecuaciones simultáneas. El problema de identificación. Mínimos cuadrados indirectos. Mínimos cuadrados en dos y tres etapas. Sistemas de ecuaciones aparentemente relacionadas (SURE). Aplicaciones con datos reales. Implementación usando software.



Métodos econométricos II: Series de Tiempo

Introducción al Análisis de Datos Temporales. Tendencias y Estacionalidad. Descomposición de series temporales. Empalme de series. Análisis gráfico. Correlación serial: Valor Esperado y Estacionariedad. Ergodicidad. Función de varianza y autocorrelación. Análisis de Correlogramas. Estrategias Básicas de Pronóstico: Variables Líderes y Asociadas. Modelo de Bass. Suavizado exponencial. Método de Holt-Winters. Ejemplos de aplicación. Modelos estocásticos básicos: Ruido Blanco. Caminos Aleatorios. Modelos Ajustados y Diagnóstico. Modelos Autoregresivos. Análisis con series macroeconómicas. Modelos de Regresión: Modelos Lineales con series temporales. Mínimos Cuadrados Generalizados. Variables Estacionales. Transformación de variables. Predicción (Forecasting) con regresión. Aplicaciones. Modelos bajo Estacionariedad: Modelos de Media Móvil, $MA(q)$. Aplicaciones. Procesos ARMA: Definición, Propiedades y aplicaciones empíricas. Modelos No Estacionarios: Modelos ARIMA estacional y no estacional. Modelos ARCH y GARCH. Aplicaciones: estimación y pronósticos. Modelos Multivariados: Regresión espúrea. Test de raíces unitarias. Cointegración. Modelo VAR. Aplicaciones.

Métodos estadísticos no paramétricos

Estimación no paramétrica de la función de densidad y de la función de regresión; test de hipótesis: Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, Spearman.

Introducción. Desafíos actuales en la minería de datos biológicos y Big-data. Proceso de minería de datos biológicos.

Datos biológicos. Tipos y características particulares. Bases de datos biológicas. Secuencias de ADN: tipos, plegado, alineación de secuencias y extracción de características. Genes, Gene Ontology (GO) y similaridad. Cómo medir distancias entre datos biológicos (distancia euclídea, correlación, similaridad semántica). Redes metabólicas y de regulación de genes.

Minería de datos. Clustering. Técnicas y modelos de clustering. Agrupamiento Jerárquico. K-medias y Partición en torno a Medoides (PAM). Mapas auto-organizativos (SOM). Ensamble de clusters. Clasificación. Redes neuronales. Máquinas de vector soporte (SVM).

Análisis estadístico de los resultados. Análisis de la varianza. Medidas de calidad en clustering. Medidas de validación externas e internas. Medidas de estabilidad en clustering.

Generación de nuevo conocimiento. Algoritmos de búsqueda de redes de relaciones. Búsqueda en profundidad y en amplitud, algoritmos evolutivos. Inferencia de redes metabólicas. Reconstrucción de redes de regulación entre genes a partir de mediciones temporales.



Modelos lineales

La esperanza condicional como el mejor predictor. El predictor óptimo lineal. Modelo de esperanza lineal homoscedástico. Modelo lineal múltiple. Discusión de los supuestos. Ecuaciones Normales. Interpretación geométrica. Test y regiones de confianza para los parámetros y para una o varias combinaciones lineales de los parámetros. Caso de rango no completo. Unicidad de la predicción. Funciones estimables y el Teorema de Gauss-Markov. Modelo con covariables aleatorias. Caso más general sin asumir normalidad de los errores. Predictor lineal óptimo. Modelo de proyección lineal. Teoría asintótica. Interpretación del modelo y ajuste bajo casos especiales de covariables: predictores cualitativos, interacción, transformaciones polinomiales de las predictoras. El modelo de análisis de la varianza (ANOVA) de un factor como caso particular. Problemas potenciales del ajuste. Algunas soluciones para la falta de linealidad o heteroscedasticidad. Estimación robusta. Técnicas de regularización. Métodos de Ridge y Lasso. Selección de modelos. Best subset regression. Método forward, backward, formatos híbridos. Criterios para elegir el modelo óptimo. Error cuadrático medio de entrenamiento y validación. Medidas de ajuste: coeficientes de determinación R^2 -ajustado, estadístico de Mallows Cp, AIC y BIC. Selección de modelos por remuestreo: validación cruzada y bootstrap. Compromiso sesgo-varianza.

Modelos lineales generalizados

Introducción a los MLG y sus componentes. Propiedades. Familia exponencial: propiedades, estimación y su relación con los MLG. Distribución muestral de los estadísticos involucrados en los MLG. Intervalos de confianza y test de hipótesis. Bondad de ajuste del modelo. Deviance. Análisis de residuos. Datos binarios: Tablas de contingencia, medidas de asociación y odds ratios. Regresión logística binaria, estimación, análisis de residuos y selección de modelo. Datos ordinales y nominales: Tablas de contingencia. Regresión logística multinomial. Estimación, análisis de residuos y selección de modelo. Enfoque de modelo de variable latente. Datos de conteo: Tablas de contingencia. Regresión de Poisson y regresión log-lineal. Estimación y análisis de residuos y selección de modelo. Casos especiales: Regresión Binomial Negativa. Extensión a modelos para respuestas repetidas o agrupadas.

Organización de las computadoras

Esquema ordenador. Organización funcional del hardware. El procesador. Arquitectura de procesadores. ISA. Procesadores mononúcleo y multinúcleo. La memoria. Organización y gestión de la memoria. Gestión de periféricos.



Problemas Inversos con aplicaciones al procesamiento de imágenes digitales

El problema de deblurring de imágenes. La función de blurring. Cómputos con matrices estructuradas. Problemas uni y bi-dimensionales. Descomposición espectral de matrices circulantes. Producto de Kronecker de matrices. Descomposición en valores singulares y análisis espectral. Introducción al filtrado espectral. Uso de SVD en reconstrucción de imágenes. Las bases DFT and DCT. La condición discreta de Picard. Regularización mediante filtrado espectral. Implementación de métodos de filtro. Análisis de errores. Métodos de elección de parámetro. Imágenes en colores y normas de suavizado. Un modelo de blurring para imágenes en colores. Regularización de Tikhonov. Deblurring por variación total. Deconvolución ciega.

Procesamiento del lenguaje natural

Introducción a modelos de lenguaje y etiquetado gramatical. Analizador léxico (tokenizer). Analizador sintáctico. Herramientas para generación de corpus de datos (scraping, procesamiento del habla). Análisis frecuencial y bolsas de palabras. Aprendizaje de representaciones vectoriales: word2vec, fasttext. Métodos basados en redes profundas: recurrentes, convolucionales y mecanismos de atención. Aplicaciones: identificación de términos y tópicos, resumen, generación de texto, sistemas de diálogo, análisis de opiniones y sentimientos.

Programación concurrente y paralela

Comunicación y sincronización. Concurrencia. Procesamiento paralelo y distribuido. Arquitecturas. Modelos de comunicación. Métricas de performance. Memoria compartida, distribuida y esquemas mixtos.

Señales y sistemas

Señales: definiciones y clasificación. Operaciones básicas con señales. Muestreo e interpolación. Concepto de ruido y relación señal a ruido. Sistemas lineales e invariantes en el tiempo. Convolución y deconvolución. Espacios de señales, bases y transformaciones lineales. Análisis de Fourier: Serie, Transformadas de tiempo continuo y discreto. Teorema de Nyquist. Transformada Z. Relación entre las distintas transformadas. Análisis y diseño de sistemas de control en el espacio de estado. Realimentación de estados. Observadores. Control óptimo.



Procesamiento digital de imágenes

Concepto de imagen digital y niveles de procesamiento. Percepción visual humana y adquisición digital de imágenes. Operaciones en el dominio espacial y frecuencial. Restauración de imágenes. Procesamiento en color. Nociones básicas de segmentación de imágenes. Morfología matemática.

Redes neuronales I

Perceptron multicapa, redes con funciones de base radial, mapas auto-organizativos, redes de Boltzmann de datos. Redes dinámicas y recurrentes: redes de Hopfield, redes de Elman y Jordan, redes con retardos temporales, retropropagación a través del tiempo. Capacidad de generalización, superficies de error, sobreentrenamiento, finalización temprana del entrenamiento.

Redes neuronales II

Aprendizaje profundo. Diferenciación automática. Redes convolucionales. Auto-codificadores. GRU, LSTM. Redes de creencia profunda. Redes generativas adversarias. Redes convolucionales en grafos. Mecanismos de atención y transformers. Redes neuronales bayesianas. Calibración en modelos profundos. Técnicas específicas de aprendizaje profundo: regularización, dropout, normalización por lotes, nuevas funciones de activación, pooling, variantes en los métodos de optimización.

Regresión y técnicas multivariadas

Regresión múltiple. Estimación. Predicción. Consideraciones del modelo de regresión: análisis de residuos, transformación de variables, variables predictoras categóricas, selección de variables. Bootstrap en modelos de regresión. Selección del modelo, validación cruzada. Métodos de clasificación. Regresión logística. Interpretación del modelo de regresión logístico ajustado. Análisis discriminante. Aprendizaje no supervisado. Análisis de componentes principales, clusters.

Sistemas operativos

Tipos de sistemas operativos. Procesos. Planificación de los procesos. Algoritmos de planificación. Problemas de concurrencia. Algoritmos. Bloqueo. Administración de memoria. Particiones fijas y variables. Memoria virtual. Paginación. Segmentación. Protección y relocalización. Administración de memorias auxiliares. Sistemas de archivos. Administración de usuarios. Seguridad.



Tecnologías de programación

Paradigmas de programación: imperativa, funcional, lógica, estructurada, orientada a objetos, dirigida por eventos, visual, orientada a aspectos. Análisis comparativo de los paradigmas. Lenguajes y aplicaciones. Tecnología de la programación para el diseño de formularios basados en disposiciones relativas, flujos, hilos de ejecución, acceso a la red. Interfases para la programación de aplicaciones.

Teoría asintótica

Teoremas básicos de teoría asintótica. Estimadores de momentos. Estimadores M y Z . Eficiencia asintótica.

Visión por computadora

Visión estereoscópica. Clasificación de imágenes. Extracción de características. Segmentación de imágenes. Registración de imágenes. Estimación de flujo óptico. Tracking. Visión computacional basada en aprendizaje.