	<p style="text-align: center;">UNIVERSIDAD DE SANTIAGO DE CHILE FACULTAD DE INGENIERÍA</p>	
---	--	--

1. PROGRAMA DE LA ASIGNATURA

CARRERA		INGENIERÍA CIVIL INFORMÁTICA					
ASIGNATURA		ANÁLISIS DE DATOS					
CÓDIGO:				T: 4	E:0	L: 2	SCT: 6
REQUISITOS		Inferencia y Modelos Estadísticos					
DICTA DEPARTAMENTO		Ingeniería Informática					
AÑO-SEMESTRE-NIVEL		4 - 2 - 8					
CATEGORIA		Obligatorio					
HORAS PRESENCIALES A LA SEMANA		6					
PERFIL DE PROFESOR		Ingeniero en computación en informática o Ingeniero con grado y especialización en análisis de datos					
VERSION	DIINF - 2015	RESOLUCIÓN FACULTAD DE INGENIERIA					
AUTOR(ES)		Max Chacón – José Luis Jara					

2. DESCRIPCIÓN DE LA ASIGNATURA

Asignatura teórico-práctica que tiene como finalidad introducir a los estudiantes en las técnicas para transformar datos en conocimiento, a través de su caracterización, separación y modelación, para Analizar o predecir un fenómeno.

3. OBJETIVOS DE APRENDIZAJE


3.1 ASOCIADOS A LA ASIGNATURA

Objetivo General:

Analizar y relacionar conjuntos de datos multivariados y/o series temporales para generar información y conocimiento, usando los modelos basados en datos que mejor expliquen el fenómeno modelado.

Objetivos Específicos:

- Representar datos multidimensionales para presentarlos de forma comprensible para su posterior análisis.
- Aplicar técnicas para identificar características en los datos que permiten su agrupamiento.
- Diseñar y evaluar clasificadores usando diferentes paradigmas de representación del conocimiento.
- Modelar linealmente relaciones entre señales y series temporales de datos, con la finalidad de realizar predicciones.

	UNIVERSIDAD DE SANTIAGO DE CHILE FACULTAD DE INGENIERÍA	
---	--	--

4. UNIDADES CONTENIDOS


UNIDAD	TÍTULO	Nº DE HORAS
1	Introducción	6
2	Análisis de Componentes Principales (ACP)	4
3	Análisis de agrupamientos	6
4	Análisis discriminante	5
5	Clasificación Bayesiana	6
6	Reglas de asociación	6
7	Árboles de decisión	8
8	Análisis de series temporales	23
	Teoría	64
	Laboratorios	34
TOTAL	17 SEMANAS	

5. CONTENIDOS DE LAS UNIDADES TEMÁTICAS

1. UNIDAD TEMÁTICA UNO: Introducción

CAPACIDADES A DESARROLLAR
<p>Conocer el proceso general de adquisición de conocimiento a partir de bases de datos.</p> <p>Comprender las diferencias entre modelamiento fenomenológico y modelamiento a partir de datos.</p> <p>Comprender los problemas de plantear hipótesis a partir de datos, sin contar con modelos fenomenológicos.</p> <p>Comprender los problemas del compromiso sesgo-varianza y el principio de parsimonia</p>

CONTENIDOS		Hrs presenciales
1.1.	Definiciones iniciales	1
1.2.	El proceso de adquisición de conocimiento (KDD)	1
1.3.	Evaluación de Hipótesis a partir de datos y selección de modelos	2
1.4.	El dilema sesgo-varianza, navaja de Occam.	2

	UNIVERSIDAD DE SANTIAGO DE CHILE FACULTAD DE INGENIERÍA	
---	--	--

2. UNIDAD TEMÁTICA DOS: Análisis de Componentes Principales (ACP)

CAPACIDADES A DESARROLLAR:
Conocer un método de reducción de dimensionalidad de los datos y usarlo para extraer información y características no explícitas de los datos.

CONTENIDOS		Hrs presenciales
2.1.	Definición de Componentes principales	
2.2.	Interpretación gráfica	1
2.3.	Análisis matemático	1
2.4.	Reducción de dimensionalidad y aplicaciones	2


3. UNIDAD TEMÁTICA TRES: Análisis de agrupamientos

CAPACIDADES A DESARROLLAR:
Conocer métodos básicos de agrupamientos jerárquicos y no jerárquicos, para agrupar datos multidimensionales, su evaluación y aplicación a problemas específicos. Analizar e interpretar los resultados de los agrupamientos.

CONTENIDOS		Hrs presenciales
3.1.	Medidas de similaridad	2
3.2.	Métodos jerárquicos	1
3.3.	K- medias, K-medias adaptivo	2
3.4.	Medidas de calidad.	1

4.

CAPACIDADES A DESARROLLAR:
Comprender los principios matemático-estadísticos de la separación de clases y las técnicas básicas del análisis discriminante de datos.

	UNIVERSIDAD DE SANTIAGO DE CHILE FACULTAD DE INGENIERÍA	
---	--	--

CONTENIDOS		Hrs presenciales
4.1.	El problema de separar grupos	
4.2.	Distancia y discriminación canónica	1
4.3.	Discriminación paramétrica	1
4.4.	Discriminación lineal	1
4.5.	Función discriminante de Fisher	1
4.5	K-vecinos más cercanos	1

5. UNIDAD TEMÁTICA CINCO: Discriminación Bayesiana

CAPACIDADES A DESARROLLAR:	
Comprender los principios de la discriminación Bayesiana. Aplicar modelos bayesianos simples (independencia condicional) a problema reales.	

CONTENIDOS		Hrs presenciales
5.1.	Discriminación bayesiana	1
5.2.	Probabilidades condicionales, Bayes e Hipótesis MAP	2
5.3.	Naive Bayes	1
5.4.	Evaluación de la clasificación	2

6. UNIDAD TEMÁTICA SEIS: Reglas de asociación

CAPACIDADES A DESARROLLAR:	
Comprender el problema de reglas de asociación, desde su problema original ("retail"). Comprender el problema combinatorio de la búsqueda de reglas y se formulación probabilista. Aplicar el concepto de reglas de asociación a diferentes problemas de la realidad.	

CONTENIDOS		Hrs presenciales
6.1.	Introducción	1
6.2.	Presentación del problema	1
6.3.	Definiciones formales y combinatoria de búsqueda	2
6.4.	Medidas de calidad	2

	UNIVERSIDAD DE SANTIAGO DE CHILE FACULTAD DE INGENIERÍA	
---	--	--

7. UNIDAD TEMÁTICA SIETE: Árboles de decisión

CAPACIDADES A DESARROLLAR:
<p>Comprender las definiciones del problema, desde la perspectiva informática y desde el punto de vista de la ganancia de la información (capacidad de canal).</p> <p>Aplicar el método como un clasificador e interpretar sus resultados al aplicarlo a problemas reales.</p> <p>Comprender la transformación de árboles a reglas.</p>

CONTENIDOS		Hrs presenciales
7.1.	Introducción	1
7.2.	Definiciones.	1
7.3.	Cálculo de Entropía y Ganancia de información	2
7.4.	Poda en árboles de decisión	2
7.5.	Transformando árboles en reglas	2

8. UNIDAD TEMÁTICA OCHO: Análisis de series temporales

CAPACIDADES A DESARROLLAR:
<p>Comprender los conceptos de la representación temporal de datos desde la perspectiva temporal, frecuencial y probabilista.</p> <p>Conocer y aplicar los modelos lineales no paramétricos y paramétricos, a secuencias temporales. Comparar modelos y determinar su eficiencia.</p>

CONTENIDOS		Hrs presenciales
8.1.	Propiedades de las series temporales	1
8.2.	Comparación de señales	1
8.3.	Estacionalidad e ergodicidad	2
8.4.	Análisis de sistemas lineales	4
8.5.	Modelos no paramétricos (función de transferencia)	2
8.6.	Modelos paramétricos (FIR, ARX, OE, ARMA)	8
8.7.	Aplicaciones	2



6. ESTRATEGIAS METODOLOGICAS

- El curso contará de clases expositivas por parte del profesor.
- Lecturas de artículos complementarios.
- Análisis de tópicos complementarios, lectura de páginas WEB, videos o películas.
- Trabajos individuales.
- Trabajos de laboratorio, donde se aplicarán los algoritmos mostrados a problemas con bases de datos reales. Se analizarán e interpretaran los resultados de los métodos estudiados.

7. EVALUACIÓN

Cátedra (60 % de nota final): NT

- Tres pruebas parciales (PEP_1 , PEP_2 , PEP_3) $\Rightarrow PEP = \overline{\{PEP_1, PEP_2, PEP_3\}}$

$$P = (PEP_1 + PEP_2 + PEP_3) / 3$$

Si ($PEP_1 \geq 4$, $PEP_2 \geq 4$, $PEP_3 \geq 4$) o ($P \geq 5$) entonces $CAT = P$

Caso contrario: Prueba Acumulativa (PA) de coeficiente 2

$$T = \text{Promedio}(\text{Top4}(PEP_1, PEP_2, PEP_3, PA, PA)).$$

Laboratorios (40 % de nota final): NL

- Seis Laboratorios (L1-L6) OBLIGATORIOS
- Nota promedio de los 6 laboratorios

Nota final:

Si $NT \geq 4,0$ y $NL \geq 4,0$ entonces APRUEBA con nota $N = 0,6 NT + 0,4 NL$
caso contrario REPRUEBA con nota $N = \min \{NT, NL\}$.



8. ASPECTOS ADMINISTRATIVOS

- El curso está regido por el reglamento general de Régimen de Estudios y el reglamento complementario de la Facultad de Ingeniería.
- Catedra y laboratorio son aprobados separadamente.
- La insistencia a una evaluación de cátedra tiene que ser debidamente justificada ante secretaria docente.
- La inasistencia a controles y la no entrega de tareas son calificadas con la nota mínima.
- Los laboratorios deben ser entregados en la fecha estipulada y no existe posibilidad de recuperación.

9. FUENTES DE INFORMACIÓN

DIRECTA

- Introducción a la Minería de datos, Hernandez, Ramirez y Ferri, Prentice Hall, 2004.
- Multivariate Statistical Methods: A primer. B.F.J. Manly, Chapman & Hall/CRC, London, 2ª Ed. 2000.
- Data Analysis, S. Brandt, Springer-Verlag. N Y, 3ª Ed. 1999.
- Intelligent Data Analysis, M. Berthold and D.J. Hand. Springer-Verlag. Heidelberg, 2ª Ed. 2003.

COMPLEMENTARIA

- <http://citeseerx.ist.psu.edu/index>
- <https://class.coursera.org/ml-005/lecture>
- <http://archive.ics.uci.edu/ml/>

10. RECURSOS ASOCIADOS

Las clases de cátedra serán realizadas en la sala de clases, que contará con notebook y proyector.

El profesor entregará el programa a los alumnos.

En la plataforma virtual el alumno encontrará:

- Programa de la asignatura del curso.
- Apuntes de clases.
- Programación de pruebas y controles.
- Planteamiento de los laboratorios.-
- Fechas de entrega de laboratorio.
- Enlaces a sitios de material complementario.