



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

GUÍA DOCENTE DE APRENDIZAJE AUTOMÁTICO

La presente guía docente corresponde a la asignatura Aprendizaje Automático (ML), aprobada para el curso lectivo 2017-2018 en Junta de Centro y publicada en su versión definitiva en la página web de la Escuela Politécnica Superior. La guía docente de ML aprobada y publicada antes del periodo de matrícula tiene el carácter de contrato con el estudiante.



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

ASIGNATURA

APRENDIZAJE AUTOMÁTICO (ML)

1.1. Código

33086 del Máster en Bioinformática y Biología Computacional

1.2. Módulo

Estadística y Análisis de Datos

1.3. Tipo

Obligatoria

1.4. Nivel

Máster

1.5. Curso

1º

1.6. Semestre

2º

1.7. Número de créditos

6 ECTS

1.8. Requisitos previos

Para un seguimiento óptimo de la asignatura se precisa tener conocimientos a nivel básico de cálculo, álgebra lineal, probabilidad y estadística, así como haber desarrollado habilidades en técnicas de programación.

1.9. Requisitos mínimos de asistencia a las sesiones presenciales

Se plantean dos itinerarios, uno con evaluación continua y otro sin ella. Los estudiantes deberán optar por uno u otro desde el principio del curso y cumplir con



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

los distintos requisitos de evaluación que conlleva cada uno de los modelos, publicados en la presente guía docente (ver apartado 4).

ITINERARIO CON EVALUACIÓN CONTINUA

La asistencia es obligatoria al menos en un 85%.

ITINERARIO SIN EVALUACIÓN CONTINUA

La asistencia es muy recomendable aunque no obligatoria.

1.10. Datos del equipo docente

Dr. Aythami Morales Moreno

Departamento de Tecnología Electrónica y de las Comunicaciones

Escuela Politécnica Superior

Despacho: C-213

Teléfono: +34 914977558

Correo electrónico: aythami.morales@uam.es

Página web: <https://atvs.ii.uam.es/atvs/>

Horario de tutorías: Petición de cita previa en clase o por correo electrónico.

Dr. Daniel Hernández Lobato

Departamento de Ingeniería Informática

Escuela Politécnica Superior

Despacho: B-346

Teléfono: +34 914972200

Correo electrónico: daniel.hernandez@uam.es

Página web: <https://dhnzl.org/>

Horario de tutorías: Petición de cita previa en clase o por correo electrónico.

Dr. Alberto Suárez González (coordinador)

Departamento de Ingeniería Informática

Escuela Politécnica Superior

Despacho: B-325

Teléfono: +34 914977531

Correo electrónico: alberto.suarez@uam.es

Página web: www.eps.uam.es/~gaa

Horario de tutorías: Petición de cita previa en clase.



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

1.11. Objetivos del curso

Esta asignatura proporciona al estudiante formación en paradigmas avanzados de aprendizaje automático. Se estudiarán algoritmos capaces de generalizar comportamientos y reconocer patrones a partir de información suministrada en forma de ejemplos.

El proceso de inducción del conocimiento conlleva seguir un flujo de trabajo realizando sucesivas iteraciones, si fuera necesario. El estudiante aprenderá a identificar qué técnicas emplear en cada una de las fases de un típico problema de aprendizaje automático: formalización del problema, identificación de las variables relevantes, pre-procesamiento de datos, construcción de modelos y validación de dichos modelos.

Las competencias básicas y generales que el estudiante adquiere en esta asignatura son:

- CG1 - Capacidad para comprender y aplicar métodos y técnicas de investigación en el ámbito de la Bioinformática.
- CG2 - Capacidad para proyectar, calcular y diseñar productos bioinformáticos.
- CG3 - Capacidad para trabajar en equipos multidisciplinares, comunicándose eficientemente y desarrollando su actividad de acuerdo con las buenas prácticas científicas.
- CG4 - Capacidad para la investigación, desarrollo e innovación, en empresas y centros tecnológicos, en el ámbito de la Bioinformática.
- CG5 - Capacidad para la aplicación de los conocimientos adquiridos y resolución de problemas en entornos nuevos o poco conocidos en el ámbito de la Bioinformática.
- CG6 - Capacidad de búsqueda, análisis y gestión de información; incluyendo la capacidad de interpretación y evaluación con un razonamiento crítico y autocrítico.
- CG7 - Capacidad de estudiar y resolver problemas biológicos y biomédicos con el soporte de herramientas computacionales.
- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida auto-dirigido o autónomo.

Las cualificaciones ubicadas en el nivel de competencias transversales que el estudiante adquirirá en esta asignatura son:

- CT1 - Capacidad para trabajar en equipo de forma colaborativa y con responsabilidad compartida en el diseño y comunicación de estrategias experimentales.
- CT2 - Capacidad de identificar fuentes de información científica solventes para fundamentar el estado de la cuestión de un problema bioinformático y poder abordar su resolución.

La competencia de tecnología específica que el estudiante adquiere en esta asignatura es:

- CE1 - Capacidad de aplicar los conocimientos de biología, matemáticas, física y estadística a la bioinformática.
- CE2 - Capacidad de aplicar métodos computacionales a la resolución de problemas en el ámbito de la investigación biomédica básica y traslacional.
- CE6 - Capacidad de explotar tecnologías avanzadas de aprendizaje automático y minería de textos.
- CE10 - Capacidad de diseñar, implementar y evaluar una solución informática para resolver necesidades en el procesamiento de datos.

Al final del semestre (objetivos generales), y de cada unidad (objetivos por tema) el estudiante deberá ser capaz de:

OBJETIVOS GENERALES	
G1	Comprender el aprendizaje automático como un proceso que permite extraer conocimiento a partir de datos.
G2	Identificar los principios en los que se basan los distintos sistemas de aprendizaje automático, así como sus ventajas y limitaciones. Conocer los métodos disponibles para su diseño y construcción.
G3	Formalizar problemas de aprendizaje automático en el ámbito de la Bioinformática y la Biología Computacional, incluyendo la identificación de variables relevantes, el pre-procesamiento de datos, la construcción de modelos y, finalmente, la validación de dichos modelos.
G4	Elegir de entre distintos modelos el apropiado para abordar un problema de aprendizaje automático dado, de acuerdo con la calidad e inteligibilidad de sus predicciones, y teniendo asimismo en cuenta los costes computacionales tanto en entrenamiento como en producción.



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

OBJETIVOS ESPECIFICOS POR TEMA	
TEMA 1. Introducción	
1.1.	Comprender las diferencias entre aprendizaje automático supervisado, no supervisado y aprendizaje por refuerzo.
1.2.	Identificar las etapas en el diseño de un sistema de aprendizaje automático.
TEMA II.- Aprendizaje no supervisado	
2.1.	Estimar densidades de probabilidad en una y varias dimensiones.
2.2.	Conocer y utilizar procedimientos para agrupar datos mediante técnicas de <i>clustering</i> .
TEMA III.- Aprendizaje supervisado	
3.1.	Entender el marco bayesiano para la toma de decisiones en sistemas probabilísticos.
3.2.	Conocer y utilizar distintos métodos, tanto básicos como avanzados, para la construcción de sistemas de predicción por inducción automáticos a partir de datos.
3.3.	Validar los modelos construidos mediante la estimación su capacidad de generalización por técnicas de validación y remuestreo (<i>bootstrap</i>).
3.4.	Realizar selección de variables.

1.12. Contenidos del programa

TEMA I. Introducción [6 horas]

- Modos de aprendizaje
 - Aprendizaje no supervisado
 - Aprendizaje supervisado
 - Aprendizaje por refuerzo
- Diseño de sistemas de aprendizaje
 - Recogida de datos
 - Selección y construcción de características
 - Preprocesamiento de datos
 - Filtrado
 - Suavizado
 - Detección de valores anómalos (*outliers*)
 - Manejo de valores no disponibles (*missing values*)
 - Tipos de atributos
 - Codificación de atributos (discretización, codificación 1 de K, etc.)
 - Reducción de la dimensión
 - Análisis de componentes: PCA, ICA, etc.
 - Representación de datos en bajas dimensiones: escalado multidimensional (MDS), SOM, mapas de difusión, etc.



TEMA II. Aprendizaje no supervisado [6 horas]

- Estimación de densidades de probabilidad
 - Histograma
 - Estimadores de Parzen
 - Modelos de mezclas
 - Mezclas de gaussianas
 - El algoritmo EM
 - Estimaciones en varias dimensiones: La maldición de la dimensión
- Análisis de conglomerados (*clustering*)
 - *Clustering* jerárquico
 - K-means
 - Fuzzy K-means

TEMA III. Aprendizaje Supervisado [30 horas]

- Introducción [2 horas]
 - Inducción automática a partir de datos
 - Problemas de clasificación y de regresión
 - Teoremas “*No free lunch*”
 - Sesgos inductivos
 - Generalización y sobre-aprendizaje
 - Medidas de error
 - Regresión: error cuadrático medio (MSE), error absoluto medio (MAE), error minimax, etc.
 - Clasificación: índice de Brier, Análisis ROC (AUC: área bajo la curva ROC, *c-index*: índice de concordancia, etc.), etc.
 - Análisis del error en términos de sesgo y varianza
 - Técnicas de validación de modelos
 - Validación cruzada con particiones aleatorias
 - Valoración cruzada con K particiones
 - Validación cruzada dejando 1 fuera (*leave-one-out*)
 - Técnicas de remuestreo (*bootstrap*, *0.632+ bootstrap*, etc.)
- Teoría de la decisión [3 horas]
 - Teorema de Bayes
 - Regla de Bayes
 - ML: Máxima verosimilitud
 - MAP: Máximo a posteriori
 - Aprendizaje Bayesiano



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

- Priors sobre modelos
- Selección de modelos
- Sistemas de clasificación básicos [5 horas]
 - Modelos generativos
 - Naïve Bayes
 - Modelos gráficos
 - Modelos discriminativos
 - Análisis discriminante: DDLA, LDA, QDA
 - Vecinos próximos
 - Árboles de Decisión (iD3, C4.5, CART, etc.)
- Combinaciones de clasificadores [3 horas]
 - Mezclas de expertos
 - Conjuntos de clasificadores
 - Bagging
 - Class-switching
 - Random forest
 - Boosting
 - Gradient boosting
 - *Wisdom of the crowds y crowdsourcing*
- Selección de atributos [3 horas]
 - Tipos de métodos para la selección de variables
 - *Filter*
 - *Wrapper*
 - *Embedded*
 - Medidas de dependencia
 - Correlación lineal
 - Información mutua
 - Covarianza y correlación de distancias
 - Métodos de selección de variables
 - Métodos basados en la información mutua: JMI, mRMR,...
 - Métodos de conjuntos
 - Estrategias para evitar el sesgo en la selección de atributos
- Modelos lineales [6 horas]
 - Regresión lineal
 - Regresión multilineal
 - Métodos incrementales
 - *Stepwise regression*
 - Regresión PCA
 - PLS: *Partial Least Squares*



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

- Métodos robustos
 - Regresión lineal con términos de regularización: *Ridge regression, lasso, elastic net*
 - Regresión lineal bayesiana
- Métodos de regresión para clasificación
 - Problemas binarios:
 - Regresión logística
 - Regresión logística bayesiana
 - Problemas multiclase:
 - Regresión *softmax* (logística multinomial)
- Métodos de núcleos [3 horas]
 - Máquinas de Vectores de Soporte
 - Regularización y minimización del riesgo estructural
 - Clasificación mediante SVM
 - Regresión mediante SVR
 - Procesos gaussianos
- Redes Neuronales [5 horas]
 - Perceptrón lineal
 - Perceptrón multicapa
 - Redes neuronales profundas
 - Redes neuronales convolucionales
 - Redes neuronales recurrentes

1.13. Referencias de consulta

Bibliografía:

- Nivel introductorio
 - “An Introduction to Statistical Learning, with Applications in R” R. G. James, D. Witten, T. Hastie, and R. Tibshirani
 - “Pattern Classification” Richard O. Duda , Peter E. Hart, David G. Stork
 - “Applied Predictive Modeling“, Kuhn and Johnson. Springer
 - “Machine Learning” Tom M. Mitchell
- Nivel avanzado
 - “Pattern Recognition and Machine Learning”, Christopher M. Bishop
 - “Machine Learning: A Probabilistic Perspective”, Kevin P. Murphy
 - “Information Theory, Inference and Learning Algorithms”, David J. C. MacKay
 - “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Trevor Hastie, Robert Tibshirani, Jerome Friedman



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

- “Computer Age Statistical Inference: Algorithms, Evidence and Data Science”, Bradley Efron, Trevor Hastie
- “Deep Learning”, Ian Goodfellow, Yoshua Bengio and Aaron Courville

2. Métodos docentes

Los métodos docentes que serán empleados en la asignatura son:

- Clases teórico-prácticas apoyadas con material multimedia
- Resolución de problemas o casos prácticos en el aula
- Aprendizaje basado en problemas
- Prácticas de programación
- Análisis crítico de la literatura científica

3. Tiempo de trabajo del estudiante

		Nº de horas	Porcentaje
Presencial	Clases teóricas	30 h	50 h (33%)
	Clases prácticas	12 h	
	Tutorías	4 h	
	Actividades en el aula	4 h	
No presencial	Estudio semanal	40 h	100 h (66%)
	Realización de actividades prácticas	40 h	
	Preparación del examen (convocatoria ordinaria)	10 h	
	Preparación del examen (convocatoria extraordinaria)	10 h	
Carga total de horas de trabajo: 25 horas x 6 ECTS		150 h	



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

4. Métodos de evaluación y porcentaje en la calificación final

SISTEMA DE EVALUACIÓN	PONDERACIÓN MÍNIMA	PONDERACIÓN MÁXIMA
Exámenes	40	70
Prácticas de laboratorio	30	50
Participación en actividades en aula	0	10

En la convocatoria ordinaria, el estudiante realizará

- 1 examen intermedio obligatorio (20 % de la calificación)
- 1 examen final obligatorio (30 % de la calificación)
- Entrega de ejercicios y prácticas (30 % de la calificación)
- 1 proyecto (20 % de la calificación)

Los estudiantes que cursen la asignatura en el itinerario de evaluación no continua deberán entregar los ejercicios y las prácticas en las fechas indicadas y realizar los exámenes obligatorios. El examen final para estos estudiantes tendrá una mayor duración y podrá incluir cuestiones sobre los ejercicios y prácticas entregadas. Podrá asimismo tener una parte oral y otra escrita.

En la convocatoria extraordinaria, el estudiante realizará

- 1 examen final obligatorio (50 % de la calificación)
- Entrega de ejercicios y prácticas (30 % de la calificación)
- 1 proyecto (20 % de la calificación)

El examen de la convocatoria extraordinaria podrá incluir cuestiones sobre los ejercicios y prácticas entregados. Podrá asimismo tener una parte oral y otra escrita.

5. Cronograma

Semana	Contenido	Horas presenciales	Horas no presenciales
1	TEMA I. Introducción [6 horas]	3 horas	5 horas
2		3 horas	5 horas
3	TEMA II. Aprendizaje no supervisado [6 horas]	3 horas	7 horas
4		3 horas	7 horas
5	TEMA III. Aprendizaje Supervisado Introducción [2 horas] Teoría de la decisión [1 hora]	3 horas	7 horas
6	Teoría de la decisión [2 horas] Sistemas de clasificación básicos [1 horas]	3 horas	7 horas



Asignatura: Aprendizaje Automático
Código: 33086
Centro: Escuela Politécnica Superior
Titulación: Máster en Bioinformática y Biología Computacional
Nivel: Máster
Tipo: Obligatoria
Nº de créditos: 6

7	Sistemas de clasificación básicos [4 horas]	3 horas	6 horas
8	Combinaciones de clasificadores [2 horas]	3 horas	6 horas
9	Combinaciones de clasificadores [1 horas] Selección de atributos [2 horas]	3 horas	7 horas
10	Selección de atributos [1 horas] Modelos lineales [5 horas]	3 horas	7 horas
11		3 horas	7 horas
12	Modelos lineales [1 hora] Métodos de núcleos [2 horas]	3 horas	7 horas
13	Métodos de núcleos [1 hora] Redes Neuronales [5 horas]	3 horas	7 horas
14		3 horas	6 horas
15	- Preparación Examen Final convocatoria ordinaria		8
	Examen Final convocatoria ordinaria	2	
16	- Preparación Examen Final convocatoria extraordinaria		7
	Examen Final convocatoria ordinaria	3	