

Sumillas de electivos, v.2.0.

Maestría de Especialización en Computer Science

4 de diciembre de 2021

El presente documento muestra las sumillas actualizadas de los cursos de la MECS. Se han adicionado los resúmenes de cada curso y organizado mejor algunos temas en algunos cursos. Tenga en cuenta que estos no son los sílabos oficiales, los cuales se tendrán a inicio de semestre, por lo cual pueden haber pequeñas diferencias entre los sílabos oficiales y las sumillas presentadas en este documento. Estas posibles diferencias no serán significativas en el contenido planificado de los cursos.

1. Especialización en Data Science

Los siguientes cursos electivos son para Data Science: Estadística Avanzada (4 créditos), Modelado para el análisis de datos (3 créditos), Big Data (3 créditos), Cloud Computing (2 créditos).

1.1. Estadística Avanzada (4 créditos)

Resumen: La Ciencia de Datos tiene sus bases en la Estadística. Es por ello que los conceptos de probabilidad y estadística son fundamentales para un estudiante que desea especializarse en Ciencia de Datos. Este curso permitirá al alumno comprender el mecanismo detrás de escena de las herramientas de modelado clave en la ciencia de datos, como los mínimos cuadrados y la regresión lineal. Además aprenderás a aplicar correctamente las herramientas estadísticas para interpretar los hallazgos de manera adecuada.

Temas:

Preliminares: probabilidad, estadística descriptiva, análisis de correlación, transformaciones de datos. Análisis estadístico: variables aleatorias, distribuciones estadísticas, muestreo a partir de distribuciones, Teorema del Límite Central, intervalos de confianza, prueba de hipótesis, P-valores. Regresión Linear y Logística: regresión lineal, valores atípicos, ajuste de parámetros, regularización; clasificación y regresión logística, balance de las clases, clasificación multiclase. Regresión bayesiana. Regresión no paramétrica.

1.2. Modelado para el análisis de datos (3 créditos)

Resumen: Un modelo estadístico es una relación matemática entre una o más variables aleatorias y otras variables no aleatorias. La aplicación de modelos estadísticos a datos no

procesados ayuda a a analizar de manera estratégica, proporcionando visualizaciones intuitivas que ayudan a identificar las relaciones entre las variables y poder hacer predicciones. Este curso desarrollará las habilidades mencionadas anteriormente, así como también aprenderá a evaluar distintas posibilidades de modelaje para un determinado escenario.

Temas:

Filosofía del modelaje: razonamiento de Occar; Bias/varianza trade off. Taxonomía de modelos, Modelos de regresión: regresión lineal, árboles de decisión. Modelos de clasificación: Support Vector Machines, clusterización. Evaluación de modelos: evaluación de clasificadores, Curvas ROC, evaluación de sistemas multiclase, evaluación de modelos de predicción de valores. Evaluación de entornos: higiene de data para evaluación, amplificación de evaluaciones pequeñas. Boosting: XGBoost. Introducción a las redes neuronales.

1.3. Big Data (3 créditos)

Resumen: El curso está diseñado para que los estudiantes profundicen su conocimiento para almacenamiento, análisis, procesamiento y gestión de grandes volúmenes de información (terabytes, petabytes e inclusive exabytes). Cada día, cada hora, cada minuto se genera gran cantidad de información la cual necesita ser procesada, almacenada, analizada en tiempo real o próximo a real. Los temas principales que se revisarán en este curso son relacionados al manejo de herramientas como hadoop, hdfs, mapreduce, yarn, spark que permiten al alumno trabajar con grandes conjuntos de datos.

Temas:

Introducción a Big Data. Características de Big Data. Ecosistemas y arquitectura de Big Data. Visión global sobre sistema de Archivos Distribuidos y Procesamiento. Infraestructura On Premise y Cloud. Aplicaciones de Big Data. Hadoop. Visión global de Hadoop. Estructura y seguridad de Hadoop. HDFS: Hadoop Distributed File System. Modelo de Programación MapReduce. Aplicaciones del uso de MapReduce. Procesamiento e ingesta de datos en lote y real time. Spark. Arquitectura de Spark. Spark RDD, DataFrame y Dataset. Spark SQL. Spark Streaming. Spark Machine. Learning Library. Spark GraphX. Procesamiento de Grafos en larga escala. Pregel: A System for Large-scale Graph Processing. Distributed GraphLab Apache Giraph is an iterative graph processing system built for high scalability.

1.4. Cloud Computing (2 créditos)

Resumen: La computación en la nube, o simplemente “nube” permite el acceso a servicios y recursos computacionales sobre demanda aparentemente infinitos, de forma similar a la electricidad. De forma general, estos recursos computacionales se encuentran en centros de datos distribuidos geográficamente, y están disponibles para los usuarios a través de internet. Este modelo de computación está sustituyendo el mecanismo tradicional de consumo de recursos “on-premises” por un nuevo método robusto, ágil, y escalable. Este curso presentará los fundamentos y las principales tecnologías que hacen posible la creación de una infraestructura de nube. En la parte práctica, estudiaremos las herramientas y frameworks esenciales tanto para la creación de una infraestructura de nube como para el diseño e implementación de aplicaciones en la nube.

Temas: Introducción a Computación en Nube. Virtualización. Tipos de virtualización. Gerenciamiento de máquinas virtuales y contenedores. Plataformas basadas en Contenedores Componentes del contenedor: procesamiento, memoria, disco, y red. Plataformas distribuidas multi-contenedor. Características y arquitectura de Kubernetes. Almacenamiento en la nube Base de datos NoSQL. Modelos de programación en la nube Programación distribuida para nube: sistemas de archivos distribuidos, y MapReduce. Tecnologías para la programación distribuida en la nube: Apache Hadoop, Apache Spark, otros. Redes en Nube Escalabilidad y Elasticidad Planificación (Scheduling). Tolerancia a Fallas, Resiliencia, Confiabilidad Conceptos. Orquestación en Nube Monitoreo en Nube Arquitectura en Nube Otros tópicos en Computación en Nube Serverless y Function as a Service. IoT.

2. Especialización en Inteligencia Artificial.

Los siguientes cursos electivos son para Inteligencia Artificial: Teoría y Algoritmos para Aprendizaje de Máquina (4 créditos), Planificación en Inteligencia Artificial (3 créditos), Modelado de incertidumbre en inteligencia artificial (3 créditos), Temas avanzados en Inteligencia Artificial (2 créditos).

2.1. Teoría y Algoritmos para Aprendizaje de Máquina (4 créditos)

Resumen: En la actualidad, con el avance de la tecnología y con el hardware cada vez más potente, se han desarrollado algoritmos basados en machine learning capaces de realizar predicciones, clasificación de datos, agrupamiento de información por similitud de características, descubrimiento de patrones; entre muchas otras tareas. En ese sentido, las aplicaciones de estos algoritmos son diversas y transversales, pudiendo utilizarse, por ejemplo, en medicina, mediante el análisis de imágenes médicas para la detección de tumores en mamografías, la predicción de covid19 utilizando rayos x y en general, en muchas tareas. El curso de aprendizaje de máquina, le permitirá desarrollar las habilidades y conocimientos necesarios para poder aplicar y crear modelos que permitan resolver diferentes problemas de la vida cotidiana.

Temas:

Introducción, ejemplos y motivación. Regresión. Regresión Lineal. Regresión no Lineal. Regularización. Capacidad del Modelo LASSO, RIDGE and ELASTIC NET. Clasificación. Regresión Logística. Máquinas de soporte vectorial (SVM). KNN (Vecinos más cercanos). Árboles de Decisión Boosting. Clustering. K-Means. Mean-Shift. DBSCAN. Gaussian Mixture Model (GMM). Agglomerative Hierarchical Clustering. Redes Neuronales. Neuronas, conexiones y pesos, función de propagación. Algoritmo Back Propagation. Optimización. Aplicaciones. Reinforcement learning. Deep Learning. Convolutional Neural Network. Recurrent Neural Network. Detección y clasificación de Objetos.

2.2. Planificación en Inteligencia Artificial (3 créditos)

Resumen: El curso introduce los conceptos fundamentales del área de Inteligencia Artificial (IA), adoptando una visión unificada a través de la noción de Agentes Inteligentes. De esta forma, la tarea de construir sistemas de IA será tratada desde el punto de vista de agentes inteligentes o agentes racionales, es decir, agentes que perciben y actúan sobre el entorno con el fin de obtener el máximo éxito. El objetivo de esta disciplina es estudiar una variedad de modelos utilizados para la planificación en IA y los métodos propuestos para resolverlos, incluidos los algoritmos clásicos y los propuestos más recientemente por la comunidad científica.

Temas:

Estrategias de búsqueda básica. Espacios de Problemas (estados, metas y operadores), solución de problemas mediante búsqueda. Agentes. Búsqueda no informada: BFS, DFS, DFS con independencia iterativa. Búsqueda bidireccional. Heurísticas y búsqueda informada: primero el mejor, generic best-first, A*, búsqueda minimax. Búsqueda avanzada: Hill-climbing,

Simulated annealing, Algoritmos genéticos. Construcción de árboles de búsqueda, espacio de búsqueda dinámico, explosión combinatoria en el espacio de búsqueda. Búsqueda en adversarios: juegos, decisiones óptimas en juegos, poda alfa-beta, juegos parcialmente observables. Satisfacción de restricciones (backtracking y métodos de búsqueda local). Agentes lógicos. Lógica de primer orden. Planeamiento clásico. Grafos de planeamiento, Planeamiento como un problema de satisfabilidad. Planeamiento multiagente. Introducción al Planeamiento probabilístico.

2.3. Modelado de incertidumbre en Inteligencia Artificial (3 créditos)

Resumen: Según una visión determinista, el estado actual y el desarrollo de la universo se decidieron en su primera etapa caótica. Hoy en día mucha gente no está de acuerdo con este punto de vista porque en la realidad la incertidumbre nos rodea. En la Inteligencia Artificial, es posible que los agentes deban manejar la incertidumbre, ya sea debido a la observabilidad parcial, no determinismo, o una combinación de los dos. Es posible que un agente nunca sepa con certeza en qué estado se encuentra o donde terminará después de una secuencia de acciones. La probabilidad proporciona una forma de resumir la incertidumbre. De esta manera el agente puede realizar predicciones probabilísticas de los resultados de la acción y, por lo tanto, seleccionar la acción con la mayor utilidad esperada. Este curso se concentra en la tarea de representar y calcular en presencia de información probabilística.

Temas:

Revisión de probabilidad básica: variables aleatorias, distribuciones de probabilidad, axiomas de probabilidad, independencia, inferencia probabilística, regla de Bayes. Redes bayesianas: definición, independencia condicional. Representaciones eficientes en distribuciones condicionales. Inferencia exacta en redes bayesianas: inferencia por enumeración, el algoritmo de eliminación de variables, complejidad de inferencia exacta. Inferencia aproximada en redes bayesianas: muestreo simple, inferencia por muestreo aleatorio (Montecarlo). Razonamiento probabilístico a lo largo del tiempo: inferencia en modelos temporales, modelos de Markov ocultos, filtros de Kalman, redes bayesianas ocultas. MDP's parcialmente observables. Introducción a la Teoría de Juegos.

2.4. Temas avanzados en Inteligencia Artificial (2 créditos)

Resumen: La Inteligencia Artificial posee muchas ramas en la actualidad, siendo que dos de las más resaltantes y clásicas son el Procesamiento de Lenguaje natural y la Visión Computacional. En este curso se introducirán los fundamentos básicos de estas dos resaltantes áreas de IA.

Temas:

Procesamiento del Lenguaje Natural. Gramaticas determinísticas y estocásticas. (algoritmos de parseo). Gramáticas libres de contexto (CFGs) y algoritmos de parseo (e.g. Cocke-Younger-Kasami CYK); CFGs probabilísticas y ponderadas. Representación del significado/Semántica: Representación de conocimiento basado en logica) Roles semánticos: Representaciones temporales, Creencias, deseos e intenciones. Metodos basados en el corpus. N-gramas y Modelos ocultos de Markov (HMMs). Suavizado y back-off. Ejemplos de uso: POS

etiquetado y morfológico. Recuperación de la información: Modelo de espacio vectorial TF e IDF; Precisión y cobertura. Extracción de información. Traducción de lenguaje. Clasificación y categorización de texto: Modelo de bolsa de palabras. Visión y percepción por computador.

Visión Computacional: Adquisición de imágenes, representación, procesamiento y propiedades. Representación de formas, reconocimiento y segmentación de objetos. Análisis de movimiento Modularidad en reconocimiento. Enfoques de reconocimiento de patrones: Algoritmos de clasificación y medidas de calidad de la clasificación. Técnicas estadísticas.

3. Rutas esperadas

Cada estudiante debe llevar 6 electivos, completando en total 18 créditos. De estos electivos, 4 son de su área, y 2 son electivos de las otras áreas, de 3 créditos cada uno.

3.1. Inteligencia Artificial

- Ciclo 2: Teoría y Algoritmos para Aprendizaje de Máquina (4), Big Data(3)
- Ciclo 3: Modelado para Análisis de Datos (3), Planificación en IA (3)
- Ciclo 4: Temas avanzados en IA (2), Modelado de Incertidumbre en IA (3)

3.2. Data Science

- Ciclo 2: Estadística Avanzada (4), Big Data (3)
- Ciclo 3: Modelado para Análisis de Datos (3), Planificación en IA (3)
- Ciclo 4: Cloud Computing (2), Modelado de Incertidumbre en IA (3)