

Planeación 2022-2023

Juan Pablo Echeagaray González

20 de agosto de 2022

Índice

| | |
|---|----------|
| 1. Ciencia de Datos Aplicada en Python | 2 |
| 1.1. Introducción a la Ciencia de Datos en Python | 2 |
| 1.1.1. Temas | 2 |
| 1.2. Visualización y Representación de Datos | 2 |
| 1.2.1. Temas | 2 |
| 1.3. Machine Learning Aplicado | 3 |
| 1.3.1. Temas | 3 |
| 1.4. Minería de Texto Aplicada | 3 |
| 1.4.1. Actividades | 3 |
| 1.5. Análisis de Redes Sociales Aplicado | 4 |
| 1.5.1. Temas | 4 |
| 1.6. Literatura recomendada | 4 |
| 2. Introducción a Bases de Datos | 4 |
| 3. Herramientas de Visualización | 4 |
| 4. Álgebra Lineal | 4 |
| 4.1. Literatura recomendada | 5 |
| 5. Probabilidad y Estadística | 5 |
| 5.1. Literatura recomendada | 5 |

1. Ciencia de Datos Aplicada en Python

Estos 5 cursos de la Universidad de Michigan introducirán a los miembros del club a la Ciencia de Datos a través del lenguaje de programación Python. Estos cursos basados en competencias se diseñaron para aquellas personas con nociones básicas de Python, que quieran aplicar técnicas estadísticas, de machine learning, visualización de datos, procesamiento de texto y análisis de redes sociales a través de librerías como `pandas`, `matplotlib`, `scikit-learn`, `nltk` y `networkx` [1].

1.1. Introducción a la Ciencia de Datos en Python

1. Duración: 3 semanas
2. Encargado: Juan Echeagaray
3. Fechas: 27 de agosto al 10 de septiembre cada sábado de 9 am a 12 pm

En este curso se introducirá al alumno a los básicos de la programación en Python. Enfocándose en lambdas, lectura y manipulación de archivos csv, y la librería `numpy`. Este curso también introducirá técnicas de manipulación y limpieza de datos a través de la librería `pandas`, utilizando las estructuras de datos `Series` y `DataFrame`. Se estudiará también como usar los métodos `groupby`, `merge`, y `pivot tables`.

Al final de este curso el alumno podrá tomar datos en forma tabular, manipularlos, limpiarlos y realizar análisis estadísticos de inferencia básicos [2].

1.1.1. Temas

- Tarea 1
- Tarea 2
- Tarea 3
- Tarea 4

1.2. Visualización y Representación de Datos

1. Duración: 3 semanas
2. Encargado: Xavier Rocabado
3. Fechas: Asincrónico, medio por definir

Se introducirá al alumno a las técnicas básicas de visualización de la información haciendo uso primordialmente de la librería `matplotlib`. El curso comenzará enseñando al alumno la diferencia entre una buena y una mala visualización, así como la interpretación visual de algunos estadísticos. Después se comenzará a practicar el uso de `matplotlib` sembrando una costumbre de seguir buenas prácticas de programación para la creación de visualizaciones, aunado a esto, el alumno explorará nuevas técnicas de visualización al implementar una desde 0.

Finalmente el alumno se dará a la tarea de plantearse una pregunta de investigación que deberá poder responder mediante el uso de una visualización [3].

1.2.1. Temas

1. Introducción a los principios de la visualización de datos
2. Gráficas básicas
3. Fundamentos de gráficos
4. Proyecto integrador

1.3. Machine Learning Aplicado

1. Duración: 3 semanas
2. Encargado: Freddy Silva
3. Fechas: 19 de octubre al 2 de noviembre cada miércoles de 3 a 6 pm

Este curso introducirá al alumno a técnicas de aprendizaje máquina aplicadas, tomando un enfoque mayor en las técnicas y métodos en vez de los conceptos estadísticos detrás de ellos. El curso comenzará con una discusión de las diferencias entre el aprendizaje máquina y la estadística descriptiva, se introducirá también la librería `scikit-learn`. Se abordará el problema de la dimensionalidad así como el del agrupamiento de datos.

Se presentarán también diversos métodos de aprendizaje supervisado con el fin de que el estudiante pueda aplicarlos haciendo uso de la librería `scikit-learn`, aprenderá a la par como evaluar los modelos que implemente. El curso concluirá con modelos más avanzados, como redes neuronales así como sus limitaciones.

Al final de este curso el alumno comprenderá las diferencias entre aprendizajes supervisados y no supervisados, identificar la técnica apropiada dado el conjunto de datos, crear una característica que necesiten, y escribir el código de Python necesario para llevar a cabo su análisis [4].

1.3.1. Temas

1. Introducción a conceptos básicos de Machine Learning
2. Modelos de aprendizaje supervisado
3. Evaluación y selección de modelos
4. Modelos de aprendizaje supervisado avanzados

1.4. Minería de Texto Aplicada

1. Duración: 3 semanas
2. Encargado: Angel Aguilar
3. Fechas: 19 de noviembre al 3 de diciembre cada sábado de 9 am a 12 pm

Este curso introducirá al alumno a técnicas básicas de minería y manipulación de texto. Comienza con el entendimiento de como Python maneja texto, la estructura del mismo desde la perspectiva humana y de una máquina, se brindará también una descripción general de la librería `nltk`. Después se diversificará el enfoque hacia el uso de `regex`, limpieza de texto y preparación del mismo para su uso en modelos de machine learning.

Al finalizar el curso se aplicarán algunas técnicas de procesamiento de lenguaje natural para demostrar como es que se logra la clasificación de texto, también se demostrará como se detectan los temas de los textos para después ser agrupados [5].

1.4.1. Actividades

- Introducción a conceptos básicos de Machine Learning
- Introducción a NLTK
- Clasificación de texto
- Modelado de temáticas y análisis semántico

1.5. Análisis de Redes Sociales Aplicado

1. Duración: 3 semanas
2. Encargado: Elías Garza
3. Fechas: 19 de noviembre al 3 de diciembre cada sábado de 9 am a 12 pm

Se introducirá al alumno al análisis de redes haciendo uso de la librería **NetworkX**. El curso comienza con una descripción de lo que es el análisis de redes, así como una motivación de por qué esta modelación podría ser óptima para algunos problemas. Después se explicarán los conceptos de conectividad y robustez. Se explorarán diferentes técnicas para medir la importancia de un nodo en la red. Se finalizará el curso con la presentación de técnicas de análisis de evolución de redes a través del tiempo, así como métodos para creación de las mismas, se abordará también el problema de la predicción de enlaces [6].

1.5.1. Temas

- Introducción a redes
- Análisis de conectividad
- Centralidad de un nodo
- Evolución de redes a través del tiempo

1.6. Literatura recomendada

- Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, Géron [7]

2. Introducción a Bases de Datos

1. Duración: Por definir
2. Encargado: Xavier Rocabado, se podrían agregar más personas
3. Fechas: Por definir

3. Herramientas de Visualización

1. Duración: Por definir
2. Encargado: Xavier Rocabado, se podrían agregar más personas
3. Fechas: Por definir

4. Álgebra Lineal

- Espacios vectoriales
- Sub-espacios
- Base, conjunto generador, espacio generado
- Independencia lineal
- Independencia geométrica
- Rango
- Ortogonalidad

- Kernel
- Mapeos lineales
- Introducción al cálculo tensorial

4.1. Literatura recomendada

- Linear Algebra and Its Applications, Gilbert Strang [8]
- Linear Algebra and Optimization for Machine Learning, Charu C. Aggarwal [9]
- Mathematics for Machine Learning, Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong [10]

5. Probabilidad y Estadística

- Concepto de probabilidad
- Probabilidad condicional
- Teorema de Bayes
- Distribuciones de probabilidad
- Teorema del límite central
- Estadística descriptiva
- Estadística multi-variada
- Medidas de tendencia central
- Medidas de dispersión

5.1. Literatura recomendada

- Probability and Statistics for Engineers and Scientists, Ronald E. Walpole [11]
- Probability and Statistics for Engineering and the Sciences, Jay L. Devore [12]
- Applied Multivariate Statistical Analysis, Richard Johnson [13]

Referencias

- [1] C. Brooks, K. Collins-Thompson, V. G. V. Vydiswaran, and D. Romero, “Applied Data Science with Python.” [Online]. Available: <https://www.coursera.org/specializations/data-science-python>
- [2] C. Brooks, “Introduction to Data Science in Python.” [Online]. Available: <https://www.coursera.org/learn/python-data-analysis?specialization=data-science-python>
- [3] —, “Applied Plotting, Charting and Data Representation in Python.” [Online]. Available: <https://www.coursera.org/learn/python-plotting?specialization=data-science-python>
- [4] K. Collins-Thompson, “Applied Machine Learning in Python.” [Online]. Available: <https://www.coursera.org/learn/python-machine-learning?specialization=data-science-python>
- [5] V. G. V. Vydiswaran, “Applied Text Mining in Python.” [Online]. Available: <https://www.coursera.org/learn/python-text-mining?specialization=data-science-python>
- [6] D. Romero, “Applied Social Network Analysis in Python.” [Online]. Available: <https://www.coursera.org/learn/python-social-network-analysis?specialization=data-science-python>
- [7] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. .O’Reilly Media, Inc.”, 2019.
- [8] G. Strang, *Linear algebra and its applications*. Belmont, CA: Thomson, Brooks/Cole, 2006.
- [9] C. C. Aggarwal, L.-F. Aggarwal, and Lagerstrom-Fife, *Linear algebra and optimization for machine learning*. Springer, 2020, vol. 156.
- [10] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for machine learning*. Cambridge University Press, 2020.
- [11] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and statistics for engineers and scientists*. Macmillan New York, 1993, vol. 5.
- [12] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Cengage learning, 2011.
- [13] R. A. Johnson, D. W. Wichern *et al.*, *Applied multivariate statistical analysis*. Pearson London, UK:, 2014, vol. 6.