



IIC2115 – Programación como Herramienta para la Ingeniería (I/2024)

Laboratorio 3

Aspectos generales

- **Objetivo:** evaluar individualmente el aprendizaje sobre modelos predictivos con sklearn, en un problema práctico con datos de estimación de precio de viviendas.
- **Lugar de entrega:** Parte 1 lunes 22/04 a las 17:30, Parte 2 domingo 28 de abril a las 23:59, ambas en el repositorio privado.
- **Formato de entrega:** archivos Python Notebook L3_1.ipynb y L3_2.ipynb con las soluciones de las partes 1 y 2 del laboratorio. Los archivos deben estar ubicados en la carpeta L3. Utilice múltiples celdas de texto y código para facilitar el trabajo del cuerpo docente. Entregas que no cumplan el formato tendrán un descuento de 0,5 pts.
- **Entregas atrasadas:** el descuento por atraso para la Parte 1 es de 1 punto cada 10 minutos o fracción. El descuento por atraso para la Parte 2 es de 1 punto por cada hora o fracción.
- **Issues:** Las discusiones en las *issues* del Syllabus que sean relevantes para el desarrollo de la evaluación, serán destacadas y se considerarán como parte de este enunciado. Así mismo, el uso de librerías externas que solucionen aspectos fundamentales del problema no podrán ser utilizadas. Solo se podrán utilizar las que han sido aprobadas en las *issues*, previa consulta de los estudiantes.
- **Entregas con errores de sintaxis y/o que generen excepciones en todas las ejecuciones** serán calificados con nota 1.0.

Descripción del problema

Considere los datos contenidos en el archivo `housing.csv`, que describen información sobre el precio y características de viviendas ubicadas en Estados Unidos. La información se encuentra agregada a nivel de distrito censal, es decir, cada fila de contiene información asociadas a las viviendas de un distrito en particular.

Si bien el archivo contiene múltiples variables, aquellas que podrá utilizar y que que capturan información para las viviendas son las siguientes:

1. **housingMedianAge**: mediana de la antigüedad de una vivienda en el distrito.
2. **totalRooms**: número total de habitaciones existentes en las viviendas de un distrito.
3. **totalBedrooms**: número total de dormitorios existentes en las viviendas de un distrito,
4. **population**: población total que reside en un distrito.
5. **households**: número total de grupos familiares que viven en un distrito (grupos de personas viviendo en la misma vivienda).
6. **medianIncome**: mediana del ingreso percibido por los grupos familiares en un distrito.
7. **medianHouseValue**: mediana del precio de una vivienda en el distrito. Esta variable debe ser considerada como la etiqueta.
8. **oceanProximity**: ubicación de la vivienda con respecto al mar.

Parte 1

- a) **Preparación de los datos**: prepare los datos para su uso posterior en los modelos predictivos, realizando todos los procesos descritos en la materia y ejercicios del curso.
- b) **Análisis visual**: analice visualmente el espacio de características, utilizando la etiqueta de los datos como color de los marcadores. Comente sobre la existencia de grupos que obedezcan o no a las etiquetas y sobre la existencia de outliers.

Parte 2

- a) **Análisis de clusters**: realice un proceso de clústering sobre el espacio de características de los datos y otro sobre los datos completos (espacio de características y etiqueta). Para ambos casos, obtenga el

número óptimo de clusters con el método del codo. Analice visualmente los resultados del clústering óptimo para ambos casos, comentando sobre sus similitudes y diferencias, enfatizando el efecto que genera la etiqueta en el proceso.

b) **Predicción de la etiqueta:** entrene modelos predictivos para los siguientes casos:

- Predicción tradicional: entrenamiento de modelos para predecir directamente la etiqueta. Pruebe al menos 3 familias distintas de modelos.
- Predicción reducida: entrenamiento de modelos sobre espacios de características reducidos, es decir, los generados por técnicas de reducción de dimensionalidad. Pruebe al menos 3 tamaños distintos de espacios de características reducidos.

En todos los esquemas puede elegir la familia de modelos que quiera. Finalmente, evalúe el rendimiento de los modelos en un conjunto de test y elija el mejor modelo, justificando su decisión.

c) **Ranking de mejores características:** elija el modelo del ítem anterior que haya sido entrenado en el espacio de características completo y que haya tenido el mejor rendimiento. Luego, realice un proceso de selección de características para construir un ranking de estas. Es decir, se espera un ordenamiento de las características, que permita identificar cuáles son más y menos informativas para el modelo predictivo seleccionado.

Corrección

Es importante que deje todas las celdas de su trabajo ejecutadas antes de subir el archivo, de lo contrario se le aplicará un descuento de 0,5 pts. al puntaje total.

Para la corrección se revisarán los procedimientos desarrollados para responder las diferentes misiones y cómo utiliza las librerías en ellas. Dado lo abierto de las misiones, se espera que las respuestas incluyan análisis y visualizaciones que permitan justificar las decisiones tomadas.

Política de Integridad Académica

Los/as estudiantes de la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile deben mantener un comportamiento acorde a la Declaración de Principios de la Universidad. En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los/as estudiantes que incurran en este tipo

de acciones se exponen a un Procedimiento Sumario. Es responsabilidad de cada estudiante conocer y respetar el documento sobre Integridad Académica publicado por la Dirección de Docencia de la Escuela de Ingeniería.

Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica. Todo trabajo presentado por un/a estudiante para los efectos de la evaluación de un curso debe ser hecho **individualmente** por el/la estudiante, **sin apoyo en material de terceros**. Por “trabajo” se entiende en general las interrogaciones escritas, las tareas de programación u otras, los trabajos de laboratorio, los proyectos, el examen, entre otros.

En particular, si un/a estudiante copia un trabajo, o si a un/a estudiante se le prueba que compró o intentó comprar un trabajo, **obtendrá nota final 1.1 en el curso** y se solicitará a la Dirección de Pregrado de la Escuela de Ingeniería que no le permita retirar el curso de la carga académica semestral.

Por “copia” se entiende incluir en el trabajo presentado como propio, partes hechas por otra persona. En caso que corresponda a “copia” a otros estudiantes, la sanción anterior se aplicará a todos los involucrados. En todos los casos, se informará a la Dirección de Pregrado de la Escuela de Ingeniería para que tome sanciones adicionales si lo estima conveniente.

También se entiende por copia extraer contenido sin modificarlo sustancialmente desde fuentes digitales como Wikipedia o mediante el uso de asistentes inteligentes como ChatGPT, Gemini o Copilot. Se entiende que una modificación sustancial involucra el análisis crítico de la información extraída y en consecuencia todas las modificaciones y mejoras que de este análisis se desprendan. Cualquiera sea el caso, el uso de fuentes bibliográficas, digitales o asistentes debe declararse de forma explícita, y debe indicarse cómo el/la estudiante mejoró la información extraída para cumplir con los objetivos de la actividad evaluativa.

Obviamente, está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, **siempre y cuando se incluya la referencia correspondiente**.

Lo anterior se entiende como complemento al Reglamento del Estudiante de la Pontificia Universidad Católica de Chile (<https://registrosacademicos.uc.cl/reglamentos/estudiantiles/>). Por ello, es posible pedir a la Universidad la aplicación de sanciones adicionales especificadas en dicho reglamento.

Compromiso del Código de Honor

Este curso suscribe el Código de Honor establecido por la Universidad, el que es vinculante. Todo trabajo evaluado en este curso debe ser propio. En caso que exista colaboración permitida con otros/as estudiantes, el trabajo deberá referenciar y atribuir correctamente dicha contribución a quien corresponda. Como estudiante es un debe conocer el Código de Honor (<https://www.uc.cl/codigo-de-honor/>).