



IIC2115 – Programación como Herramienta para la Ingeniería (II/2021)

## Taller 4b

### Objetivos

- Aplicar los contenidos de modelos predictivos basados en *Machine Learning*.

### Entrega

- **Lenguaje a utilizar:** Python 3.6 o superior
- **Lugar:** repositorio privado en GitHub. Recuerde incluir todo en una carpeta de nombre **T4b**.
- **Entrega:** lunes 7 de junio a las 16:50 hrs.
- **Formato de entrega:** archivo python notebook (**T4b.ipynb**) y archivo python (**T4b.py**) con la solución de este enunciado. Los archivos deben estar ubicados en la carpeta **T4b**. No se debe subir ningún otro archivo a la carpeta. Utilice múltiples celdas de texto y código para facilitar la revisión de su programa.
- **NO SE ADMITEN ENTREGAS FUERA DE PLAZO**
- Entregas con errores de sintaxis y/o que generen excepciones serán calificadas con nota 1.0.

## Descripción del problema

Considere el conjunto de datos almacenado en el archivo `quakes.csv`, que contiene información sobre sismos ocurridos cerca de un país desconocido, durante el año 1964. Para cada sismo registrado en el archivo, se tienen los siguientes datos:

- Lat: latitud en donde ocurrió el sismo.
- Long: longitud en donde ocurrió el sismo.
- Depth: profundidad del sismo.
- Mag: magnitud del sismo (en escala de magnitud de momento).
- Stations: cantidad de estaciones que reportaron el sismo.

En base a toda esta información, complete las misiones indicadas a continuación.

### IMPORTANTE

Recuerde codificar numéricamente los valores de las columnas categóricas y normalizar las numéricas (esto último siempre y cuando no se trate de las columnas que serán objetivo de la predicción). Sea cuidadoso con el momento en que codifica y normaliza los valores (antes o después de crear los conjuntos de entrenamiento y prueba).

### Misión 1: predicción de variables numéricas

Utilizando solo registros que no tengan valores faltantes y una partición entrenamiento/test aleatoria de 70/30, construya al menos tres modelos predictivos que permitan inferir la magnitud de un sismo, a partir del resto de las variables. Usando como métrica de rendimiento el *error cuadrático medio*, cuantifique el nivel de sobreentrenamiento de cada uno de los modelos entrenados y seleccione el mejor en base a esto. Finalmente, utilice el modelo seleccionado para completar los valores faltantes de la columna **Mag**.

### Misión 2: normalización

Repita la misión anterior, esta vez utilizando como columna objetivo el logaritmo natural de la variable **Mag**. Comente los resultados, destacando: i) diferencias de rendimiento con la misión anterior, ii) cuál es el efecto de la transformación logarítmica, iii) cuándo tiene sentido aplicar la transformación logarítmica.

### Misión 3: comparación

Repita nuevamente la Misión 1, corrigiendo o eliminando los valores extremos de la variable **Mag**. Comente los resultados, destacando: i) diferencias en rendimiento con las misiones anteriores, ii) como se compara el efecto de este preprocesamiento con el de la transformación logarítmica.

### Misión 4: complejidad

Entrene alguno de los modelos disponibles en **scikit-learn** en alguna de las versiones de los datos de las misiones anteriores, esta vez modificando la complejidad del modelo (en base a 1 solo hiperparámetro) hasta encontrar aquel con el mejor rendimiento. Compare con los resultados obtenidos en las misiones anteriores.

## Corrección

La corrección de este taller se basará en lo adecuado de los mecanismos utilizados para realizar cada misión. En otras palabras, no existe *a priori* un resultado correcto para cada misión, por lo que cada misión se corregirá en base a lo adecuado y justificado que se encuentre el análisis. Cualquier supuesto que se haga para completar las misiones debe quedar claramente indicado.

**ULTRA IMPORTANTE:** todas las celdas utilizadas deben estar ejecutadas al momento de entregar el taller, de modo que las salidas generadas sean visibles. En caso de no cumplir esto, **SU TALLER NO SERÁ CORREGIDO**.

## Objetivo de participación

Para verificar la participación durante la clase, debe completar la Misión 1.

## Política de Integridad Académica

*“Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad.”*

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones,

el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por “copia” o “plagio” se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.