



IIC2115 – Programación como Herramienta para la Ingeniería (II/2017)

Actividad 12

Objetivos

- Introducir diferentes herramientas de análisis de datos en Python.

Entrega

- **Lenguaje a utilizar:** Python
- **Lugar:** GitHub
- **Hora:** 17:30

Instrucciones

Nivel 1: Bastián

El *dataset* Anscombe's quartet (adjunto al enunciado), consiste en 4 conjuntos de una serie de pares (x, y) . Laika, la perrita favorita de todos, te recomienda imprimir algunas métricas interesantes, para obtener mayor conocimiento de estos datos. En particular, Laika quiere ver las siguientes métricas, que pueden ser calculadas utilizando la librería **pandas**.

1. Media.
2. Desviación estándar.
3. Máximo.
4. Mínimo.
5. Correlación entre las variables x e y .
6. Kurtosis.

Es importante notar que se esperan los valores de estas métricas por **Set**, es decir, debes obtener las métricas para el set A, para el B, para el C y para el D, separadamente.

Nivel 2: Laika

¡Sorpresa! La media, la desviación estándar y la correlación entre variables de los cuatro **Sets** son las mismas. Sin embargo, ¿es posible afirmar que todos tienen comportamientos exactamente iguales? Una gran idea para responder esta pregunta es visualizar los datos. Por lo tanto, para completar este nivel deberás generar cuatro gráficos de dispersión (*scatter plots*, uno por **Set**), donde se evidencie la relación (x, y) de los pares que corresponden a cada **Set**. Para esto, Laika sugiere que utilices su librería favorita: **matplotlib**.

Nivel 3: Hugo

¡Felicitaciones! Has llegado al último nivel de esta actividad. Ahora utilizaremos el *dataset* sobre diabetes, incluido junto a `sklearn`, para practicar regresión lineal. Los pasos a seguir en este nivel son:

1. Carga el *dataset*.
2. Usa `.DESCR` para imprimir una descripción detallada del *dataset*.
3. Define X e y , donde X son los datos que el modelo debe aprender (*data*) e y es la variable a predecir (*target*). Para este paso, se sugiere fuertemente que explores las llaves del objeto *dictionary-like* que retorna `load_diabetes`.
4. Separa los datos en entrenamiento y *testing*. Para esto, puedes utilizar `train_test_split` de `sklearn`.
5. Ajusta la regresión lineal a los datos de entrenamiento, con la ayuda de `sklearn`.
6. Usando `.coef_`, revisa los coeficientes de la regresión y responde: ¿Cómo se interpreta el coeficiente correspondiente al índice de masa corporal (*bmi*, por su sigla en inglés)?
7. Predice la medida de progresión de la enfermedad (el *target*), usando para eso los datos de *testing*.
8. Realiza un gráfico donde se aprecien los valores reales contrastados con los predichos.

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

“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.