

**UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS**  
**CIENCIAS DE LA COMPUTACIÓN**

**MACHINE LEARNING**  
**Laboratorio Regresión Lineal**  
**(Primer Semestre del 2019)**

Objetivos de aprendizaje:

- Entrenar modelos de regresión lineal.
- 

## 1. Actividad en Weka

**Predicción de precio de venta de una casa (15 minutos)**. Se desea predecir el valor de venta de una determinada casa. Los datos de determinadas ventas se encuentran en el archivo `houses.arff`. Se le pide que entrene un modelo usando el algoritmo `LinearRegression` usando los parámetros por defecto de algoritmo. Una vez obtenido el modelo, realice una predicción utilizando los siguientes datos:

- `House size = 3198`
- `Lot size = 9669`
- `Bedrooms = 5`
- `Granite = 1`
- `UpgradedBathroom? = 1`

### Preguntas de discusión

- Describa al modelo. ¿Es un buen modelo? Verifique el valor del coeficiente de correlación.
- Es factible ejecutar la validación cruzada en este conjunto de datos. ¿Por qué?
- ¿Qué pasa cuando en lugar de usar el algoritmo `LinearRegression` usamos el algoritmo `SimpleLinearRegression`? ¿Qué diferencias encuentra entre ambos algoritmos?

**Predicción de rendimiento de Millas por Galón (15 minutos)**. Se desea predecir el cuántas millas rinde un galón dada ciertas condiciones del carro como el número de cilindros, la cilindrada, los caballos de fuerza, el peso, la aceleración, el modelo y su origen. Los datos de determinadas vehículos se encuentran en el archivo `autoMpg.arff`. Se le pide que entrene un modelo de regresión lineal.

### Preguntas de discusión

- Describa al modelo. ¿Es un buen modelo? Verifique el valor del coeficiente de correlación.
- Es factible ejecutar la validación cruzada en este conjunto de datos. ¿Por qué?
- ¿Qué pasa cuando en lugar de usar el algoritmo `LinearRegression` usando el algoritmo `SimpleLinearRegression`?

## 2. Actividad en Python

**Predicción de la eficiencia energética (30 minutos)** En el archivo `Energy_efficiency.csv` se almacena un conjunto de datos en donde se registran ciertos parámetros de construcción, la carga de calefacción y la carga de refrigeración. Los parámetros de construcción están representados por los atributos **X1** (*Relative Compactness*), **X2** (*Surface Area*), **X3** (*Wall Area*), **X4** (*Roof Area*), **X5** (*Overall Height*), **X6** (*Orientation*), **X7** (*Glazing Area*) y **X8** (*Glazing Area Distribution*). La **carga de calefacción** está representado por el atributo **y1** (*Heating Load*) y la **carga de refrigeración** por el atributo **y2** (*Cooling Load*).

La eficiencia energética se puede definir en función de los valores de la carga de calefacción y la carga de refrigeración. Se desea saber si es que es posible definir la eficiencia energética en términos de los parámetros de construcción. Se desea encontrar el modelo de regresión lineal más óptimo para predecir el conjunto de datos. La descripción completa de este conjunto de datos la puede encontrar en la URL <https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>.

Use la técnica de la validación cruzada. Incluya dentro de los algoritmos a probar el algoritmo `LinearRegression`. Vea el siguiente programa para que sepa la librería a importar y los comandos a usar para el entrenamiento.

```
1 from sklearn.linear_model import LinearRegression
2
3 linealregression = LinearRegression()
4 linealregression.fit(X_entrenamiento, y_entrenamiento)
```

### Preguntas de discusión

- ¿Cuál es el coeficiente de correlación del modelo?
- ¿Qué atributos seleccionó para entrenar el modelo? ¿Cómo justifica su elección?
- ¿El valor de este modelo algorítmico es consistente con el modelo obtenido en Weka?
- ¿Existe alguna diferencia si es que normalizamos los atributos?

### Normalización de atributos en Weka

Para normalizar los atributos en Weka siga el siguiente procedimiento:

- Diríjase a la pestaña **Preprocess**.
- Presione el botón **Choose** que se encuentra en el panel **Filter**.
- Seleccione de la lista el filtro **Normalize** que se encuentra en la categoría de filtros **Unsupervised**, del tipo **Attribute**.
- Presione el botón **Apply** que se encuentra en el panel **Filter**.

Monterrico, 13 de mayo de 2019.