

TALLER DE PROCESAMIENTO DE SEÑALES


1er Cuatrimestre 2026 - Trabajo Práctico Nº 2 - Regresión Polinómica

Se desea predecir la esperanza de vida de un país. Para ello se cuenta con datos del gasto público en salud per-cápita (adaptado por inflación) para diferentes países.


(a) *Creación de la base de datos:*

1. Cargar la base de datos utilizando `load_dataset('healthexp')` (seaborn).
2. Graficar un `scatter` de la esperanza de vida en función del gasto público indicando con un color distinto los diferentes países.
3. Definir el conjunto de datos de entrenamiento a partir de la información anterior al 2008, el de validación con los datos entre el 2008 y el 2015, y el de testeo con los posteriores al 2015.

(b) *Preprocesamiento*

1. Utilizar el comando `OneHotEncoder` (sklearn) para codificar la información del país al que corresponde el dato como variable *one-hot*. Explicar su funcionamiento.
2. Utilice el comando `PolynomialFeatures` (sklearn) para crear un mapa polinómico de orden 3 sobre la variable de gasto público en salud per-cápita.
3. Combinar y posteriormente normalizar las salidas de las operaciones anteriores. : Herramientas recomendadas son `ColumnTransformer` y `StandardScaler` (sklearn).

(c) *Regresión Polinómica:*

1. Utilizar `Pipeline` (sklearn) para combinar las operaciones anteriores con una regresión lineal regularizada. : Se recomienda utilizar `RegressionLineal` (sklearn).
2. ¿Cuántos parámetros tiene el modelo? Explicar por qué.
3. Utilizar `set_config(display="diagram")` (sklearn) para hacer un diagrama del sistema definido.
4. Entrenar el sistema descrito anteriormente.
5. Reportar el error cuadrático medio de entrenamiento y validación.
6. Graficar la regresión obtenida por país, sobre el `scatter` de los datos.

(d) *Regularización:*

1. Utilizar `Ridge` (sklearn) para entrenar una regresión lineal regularizada para diferentes valores de $0 < \lambda < 1$.
2. Graficar el error cuadrático medio de entrenamiento y validación en función del hiperparámetro λ .
3. Reportar el λ que minimiza el error cuadrático medio de validación.
4. Reportar el error cuadrático medio de testeo para el λ obtenido anteriormente.
5. Graficar la regresión obtenida por país, sobre el `scatter` de los datos.

(e) *Teoría:* Se desea justificar las expresiones relacionadas con el riesgo empírico regularizado. Tener en cuenta la presencia del sesgo en el modelo, pero no en la regularización

1. Mostrar la expresión del gradiente del riesgo.
2. Hallar una solución matricial para los parámetros.