

Aprendizaje Automático

Examen

Fecha: 16-10-2025

Ejercicio 1

Utilice el dataset **Boston Housing** disponible en scikit-learn. El objetivo es predecir el valor mediano de las viviendas. Compare el desempeño de los siguientes modelos:

- a) Modelo lineal gaussiano
- b) Modelo lineal con regularización Elastic Net
- c) Modelo lineal con componentes principales (PCA)
- d) Máquinas de Vector de Soporte para regresión (SVR)
- e) Árbol de decisión para regresión

Tareas específicas:

1. Cargue los datos y realice un análisis exploratorio básico

```
from sklearn.datasets import load_boston
boston = load_boston()
X, y = boston.data, boston.target
```

2. De ser el caso, realice búsqueda de hiperparámetros mediante validación cruzada usando 7 folds.
3. Evalúe el desempeño usando MAE y RMSE en 50 conjuntos de prueba y luego halle la mediana y desviación estandar. Para obtener cada conjunto, divida los datos en entrenamiento (70 %) y prueba (30 %).
4. Compare todos los modelos usando una tabla resumen
5. Justifique cuál modelo elegiría y por qué.

Ejercicio 2

Utilice el dataset **Stroke Prediction** disponible en Kaggle, el mismo que se puede descargar desde: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>.

La pregunta de investigación es:

¿Cuáles son los principales factores de riesgo para accidente cerebrovascular y cómo podemos predecir eficientemente pacientes en riesgo?

Tareas específicas:

1. Cargue y realice un análisis exploratorio del dataset.
2. Diseñe **dos estrategias de modelado** distintas:
 - Estrategia A: Enfoque inferencial para identificar factores de riesgo
 - Estrategia B: Enfoque predictivo para clasificación de alto riesgo