

Elección de técnica de escalamiento según el algoritmo

Contexto:

El tipo de algoritmo de Machine Learning define si es necesario escalar todos los datos y qué técnica conviene. Esta actividad ayuda a decidir con criterio técnico.

Consigna:

Analiza distintos algoritmos de ML y decide si requieren escalamiento, y en ese caso, cuál es la técnica más adecuada (Min-Max o Z-score). Justifica tus decisiones.

Paso a paso:

1. El docente presentará una lista de 5 algoritmos:
 - K-Nearest Neighbors (KNN)
 - Árbol de Decisión (Decision Tree)
 - Regresión Lineal
 - Support Vector Machine (SVM)
 - Random Forest
2. Para cada algoritmo responde:
 - ¿Es sensible a la escala de las variables?
 - ¿Necesita escalamiento? ¿Cuál conviene más?
3. Discute las decisiones en grupos pequeños.
4. Arma una tabla resumen con tus conclusiones.

1. K-Nearest Neighbors (KNN)

- **¿Es sensible a la escala?**
Sí, porque KNN se basa en cálculos de distancia (normalmente Euclidiana) entre puntos de datos. Si las variables tienen diferentes escalas, las de mayor rango dominarán la distancia.

- **¿Necesita escalamiento?**
Sí, es muy recomendable escalar.
- **¿Cuál conviene más?**
Min-Max o Z-score (normalización o estandarización) funcionan bien. La elección puede depender de los datos:
 - Min-Max: si quieres que los datos estén en un rango definido (0 a 1), útil si el algoritmo espera inputs acotados.
 - Z-score: si hay valores atípicos, estandarizar puede ayudar.

2. Árbol de Decisión (Decision Tree)

- **¿Es sensible a la escala?**
No, porque las decisiones se basan en umbrales sobre valores individuales de cada variable, no en distancias.
- **¿Necesita escalamiento?**
No es necesario.

3. Regresión Lineal

- **¿Es sensible a la escala?**
En teoría no afecta la capacidad predictiva, pero la escala sí impacta la interpretación de coeficientes y la convergencia del algoritmo cuando se usa regularización.
- **¿Necesita escalamiento?**
Sí, si se usan técnicas de regularización (Ridge, Lasso), el escalamiento es importante para que todas las variables tengan la misma importancia en la penalización.
- **¿Cuál conviene más?**
Z-score (Estandarización), porque centra y escala, facilitando la convergencia y comparación de coeficientes.

4. Support Vector Machine (SVM)

- **¿Es sensible a la escala?**
Sí, SVM busca maximizar márgenes y usa distancias en su cálculo, por lo que la escala afecta el resultado.

- **¿Necesita escalamiento?**
Sí, es fundamental.
- **¿Cuál conviene más?**
Z-score es la técnica más común, ya que el SVM suele funcionar mejor con datos estandarizados (media 0, desviación 1).

5. Random Forest

- **¿Es sensible a la escala?**
No, al igual que los árboles, Random Forest divide según valores individuales, sin basarse en distancias.
- **¿Necesita escalamiento?**
No es necesario.

Algoritmo	Sensible a la escala	Necesita escalamiento	Técnica recomendada	Justificación
KNN	Sí	Sí	Min-Max o Z-score	Se basa en distancias; evita que una variable domine.
Árbol de Decisión	No	No	No aplica	Usa umbrales; no afecta la escala.
Regresión Lineal	Parcialmente	Sí (con regularización)	Z-score	Mejora convergencia y comparación de coeficientes.
SVM	Sí	Sí	Z-score	Márgenes y distancias requieren variables en misma escala.
Random Forest	No	No	No aplica	Basado en árboles; insensible a la escala.