

CONSIDERACIONES EJERCICIO REGRESIÓN SUPERHEROES

MÉTRICAS A INTERPRETAR

1. RMSE (Root Mean Squared Error)

- Qué es: Promedio de los errores de predicción

Interpretación:

Valor más bajo = mejor predicción

Si variable dependiente va de 0-100, un RMSE de 15 significa error promedio de ± 15 puntos

Comparar entre conjuntos:

$\text{RMSE}(\text{train}) < \text{RMSE}(\text{val}) < \text{RMSE}(\text{test}) \rightarrow \text{Normal}$

$\text{RMSE}(\text{train}) \ll \text{RMSE}(\text{val}/\text{test}) \rightarrow \text{Overfitting}$

$\text{RMSE}(\text{train}) \approx \text{RMSE}(\text{val}) \approx \text{RMSE}(\text{test}) \rightarrow \text{Modelo generaliza bien}$

2. MAE (Mean Absolute Error)

Qué es: Error promedio sin penalizar tanto los outliers

Interpretación:

Más robusto que RMSE a valores extremos

Si MAE = 10, en promedio te equivocas 10 puntos

3. R^2 (Coeficiente de determinación)

Qué es: Proporción de varianza explicada (0 a 1)

Interpretación:

$R^2 = 0.85 \rightarrow$ El modelo explica 85% de la variabilidad

$R^2 = 0.50 \rightarrow$ Explica 50% (moderado)

$R^2 = 0.20 \rightarrow$ Explica poco, modelo débil

$R^2 < 0 \rightarrow$ Modelo peor que simplemente predecir la media

4. Número de componentes PCA

Qué es: Cuántos componentes principales se necesitan para capturar $\geq 95\%$ varianza

Interpretación:

Si $n_{\text{comp}} = 3$ de 5 \rightarrow Lograste reducción dimensional efectiva

Si $n_{\text{comp}} = 5$ de 5 \rightarrow No hubo reducción (todas las variables son importantes)

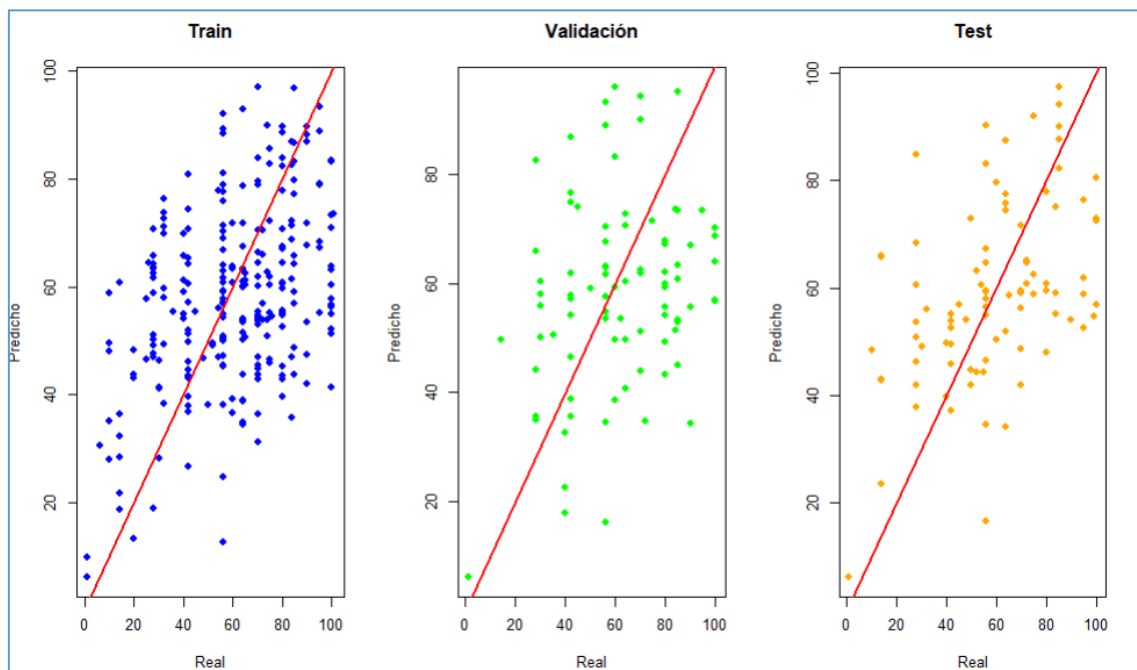
SEÑALES DE ALARMA:

```
OVERFITTING:
Train RMSE: 5    → Predice muy bien en train
Val RMSE: 25    → Predice mal en val
Test RMSE: 28    → Predice mal en test

BUEN MODELO:
Train RMSE: 15
Val RMSE: 17
Test RMSE: 16

UNDERFITTING:
Train RMSE: 30    → Predice mal en todos
Val RMSE: 32
Test RMSE: 31
```

Gráfica del modelo con variable COMBAT objetivo



Cada gráfico compara:

- **Eje X (Real):** Valor verdadero de Combat
- **Eje Y (Predicho):** Valor que predijo el modelo
- **Línea roja diagonal:** Predicción perfecta (si predicho = real)
- **Si los puntos estuvieran todos pegados a la línea roja:** → $R^2 = 1.0$ (perfecto)
- **En el modelo (puntos cerca, pero dispersos):** → $R^2 = 0.66$ (bueno)
- **Si los puntos estuvieran por todo el gráfico sin patrón:** → $R^2 = 0$ (malo)
- El 34% de varianza no explicada ($1 - 0.66 = 0.34$) se ve visualmente como:
 - La distancia vertical de cada punto a la línea roja
 - Mientras más lejos del rojo, más error
 - Outliers evidentes