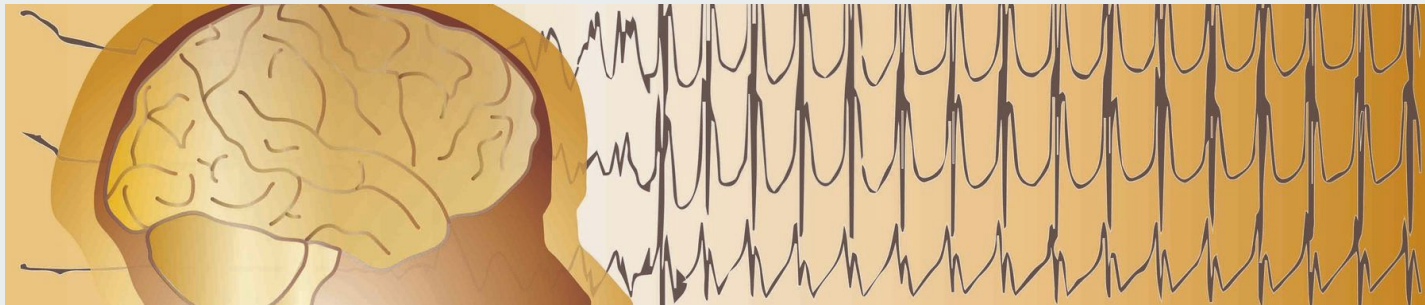


Reconocimiento de Crisis Epilépticas vía electroencefalogramas (EEG)

Técnicas de Reconocimiento de Patrones
Jaime Benhamu Oaknin y Rocío Jiménez Villén



Contexto

Epilepsia: Trastorno neurológico causado por activación anormal de neuronas.

Electroencefalograma (EEG): Registra la actividad eléctrica del cerebro a través de ondas de diferentes frecuencias.

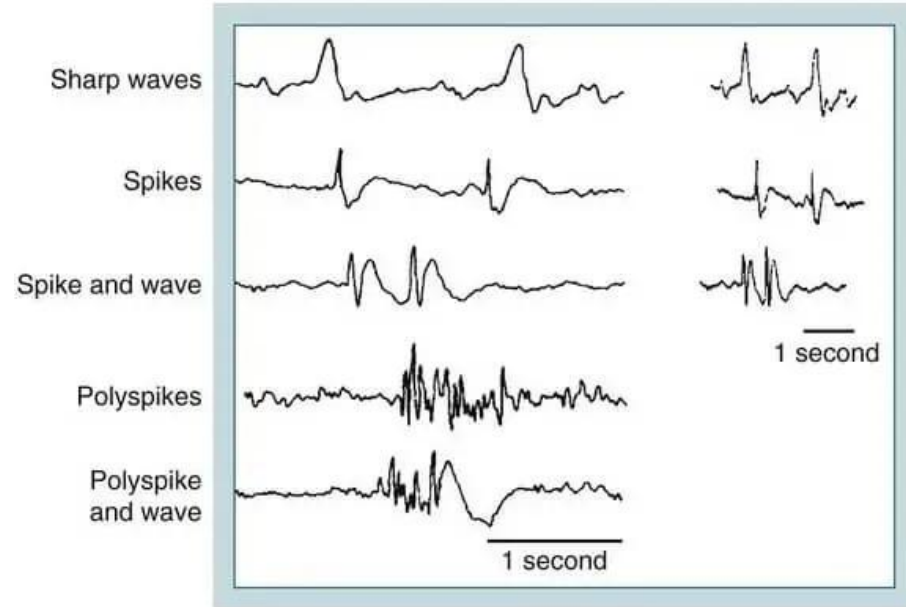
- ♦ **EEG en personas sanas:**

- Neuronas trabajan de forma síncrona.
- Cada estado cerebral tiene ondas características.

- ♦ **EEG en epilepsia:**

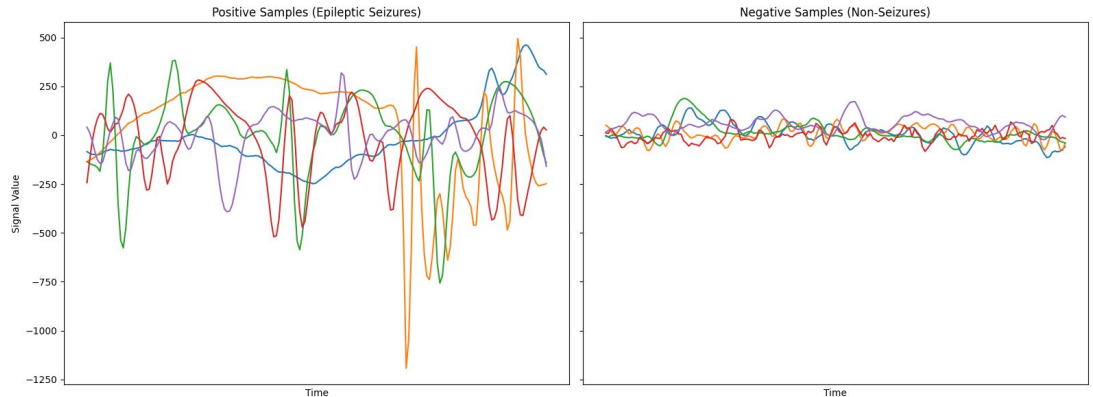
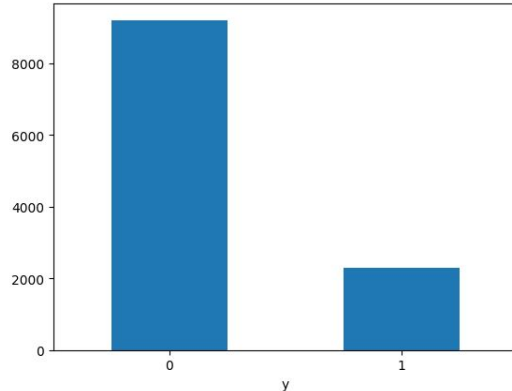
- Patrones anormales, como picos o descargas de alta frecuencia.
- Actividad neuronal desorganizada.

Patrones comunes en EEG que registran epilepsia



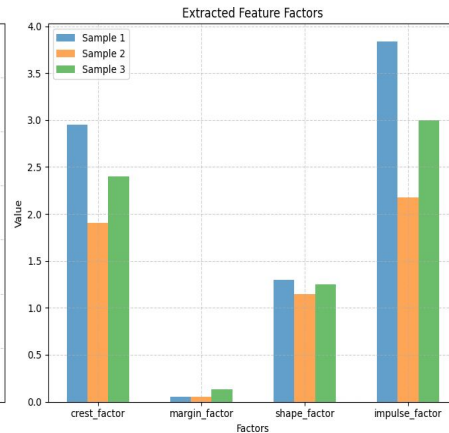
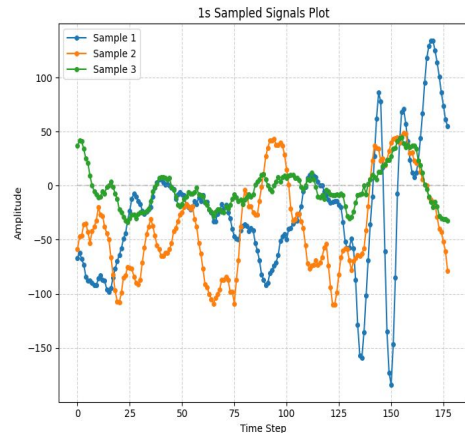
Dataset: Epileptic Seizure Recognition

- Extraído de Kaggle, construido a partir del conjunto de datos de UCI Machine Learning.
- Datos EEG de 500 individuos, con registros de 23 segundos de duración.
- Cada registro de 23 segundos se divide en ventanas de 1 segundo, generando 23 muestras por individuo.
- Total de muestras: $500 \times 23 = 11,500$
- Cada muestra de EEG está discretizada en 178 timesteps (puntos temporales).
- Tarea de clasificación binaria: distinguir entre crisis epilépticas (1) y no crisis epilépticas (0).



Extraccion de Características: Time domain

	mean	std	var	min	max	skew	kurtosis	rms	zero_crossings	abs_max	crest_factor	margin_factor	shape_factor	impulse_factor
0	-16.910112	95.980947	9212.342157	-281.0	229.0	-0.200327	0.067411	97.193309	23.0	281.0	2.891146	0.030503	1.256019	3.631334
1	28.112360	473.166815	223886.834762	-1716.0	513.0	-1.511088	1.341888	472.672559	8.0	1716.0	3.630420	0.007665	1.258349	4.568335
2	-44.044944	44.311025	1963.466895	-126.0	80.0	0.494485	-0.240414	62.389048	12.0	126.0	2.019585	0.064172	1.157038	2.336737
3	-68.910112	15.968642	254.997524	-105.0	-22.0	0.367125	0.212160	70.726010	0.0	105.0	1.484602	0.411769	1.026352	1.523724
4	-6.651685	38.802149	1505.606805	-103.0	78.0	-0.462741	-0.250516	39.260581	32.0	103.0	2.623497	0.068411	1.269692	3.331032

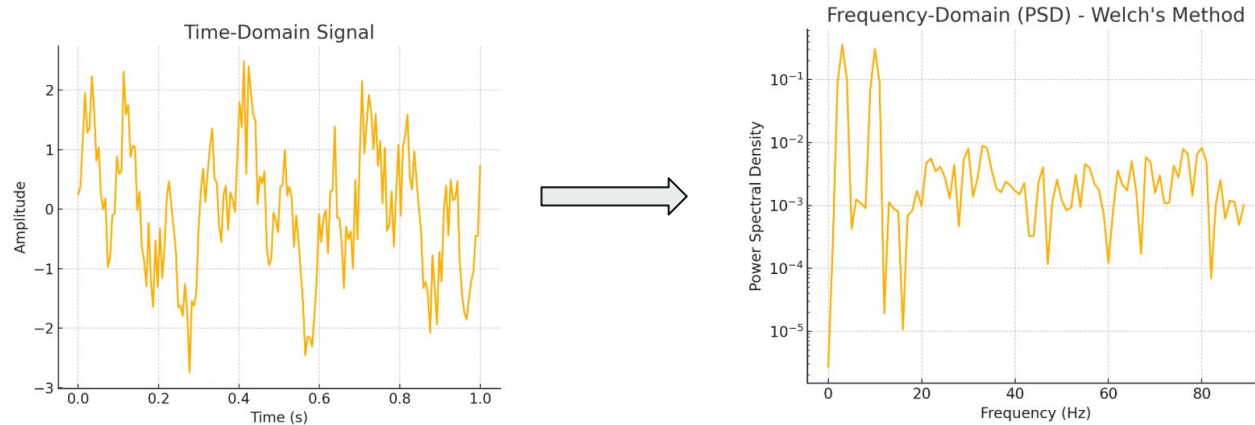


$$C(y_n) = \frac{\max |y_n|}{RMS(y_n)} \quad \text{where } RMS(y_n) = \sqrt{\frac{1}{N} \sum_{n=1}^N y_n^2}$$

$$M = \frac{\max |y_n|}{\text{Var}(y_n)}, \quad \text{where } \text{Var}(y_n) = \frac{1}{N} \sum_{n=1}^N (y_n - \bar{y})^2$$

$$S(y_n) = \frac{RMS}{MAV}, \quad MAV = \frac{1}{N} \sum_{n=1}^N |y_n|$$

Extracción de características: Frequency Domain



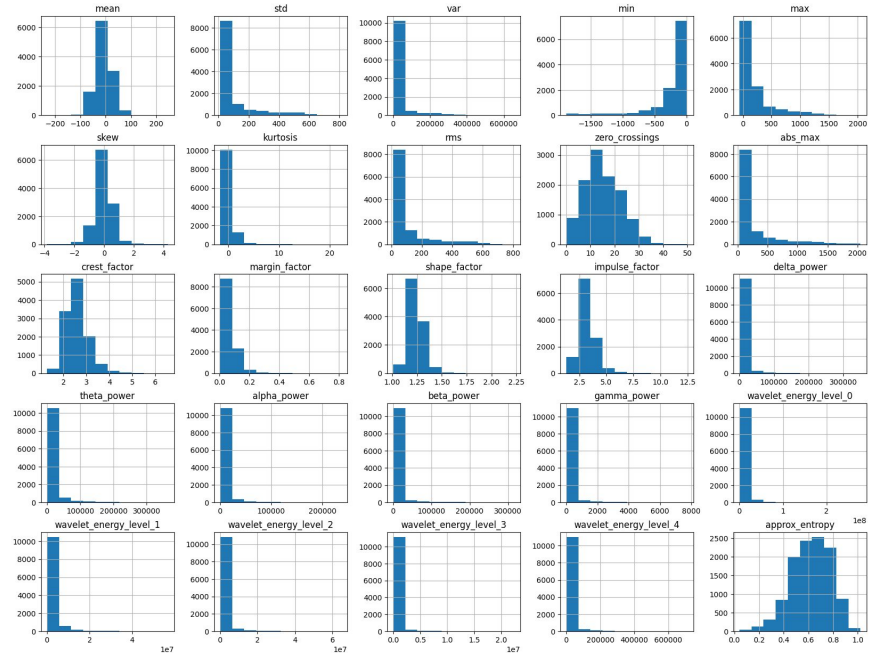
Power Band	Frequency Range (Hz)	Description
Delta (δ)	0.5 – 4 Hz	Associated with deep sleep and restorative processes. Abnormally high delta waves during wakefulness may indicate brain injuries or epilepsy.
Theta (θ)	4 – 7 Hz	Linked to drowsiness, light sleep, and memory processing. Increased theta activity may occur before seizure onset in epilepsy patients.
Alpha (α)	8 – 12 Hz	Found in relaxed wakefulness, particularly with closed eyes. Alpha suppression is observed during seizures and cognitive engagement.
Beta (β)	13 – 30 Hz	Associated with active thinking, problem-solving, and alertness. Abnormal bursts of beta waves can be seen in focal epilepsy.
Gamma (γ)	30 – 100 Hz	Linked to high-level cognitive functions and sensory processing. Increased gamma activity is sometimes observed during seizure episodes.

Selección de Características

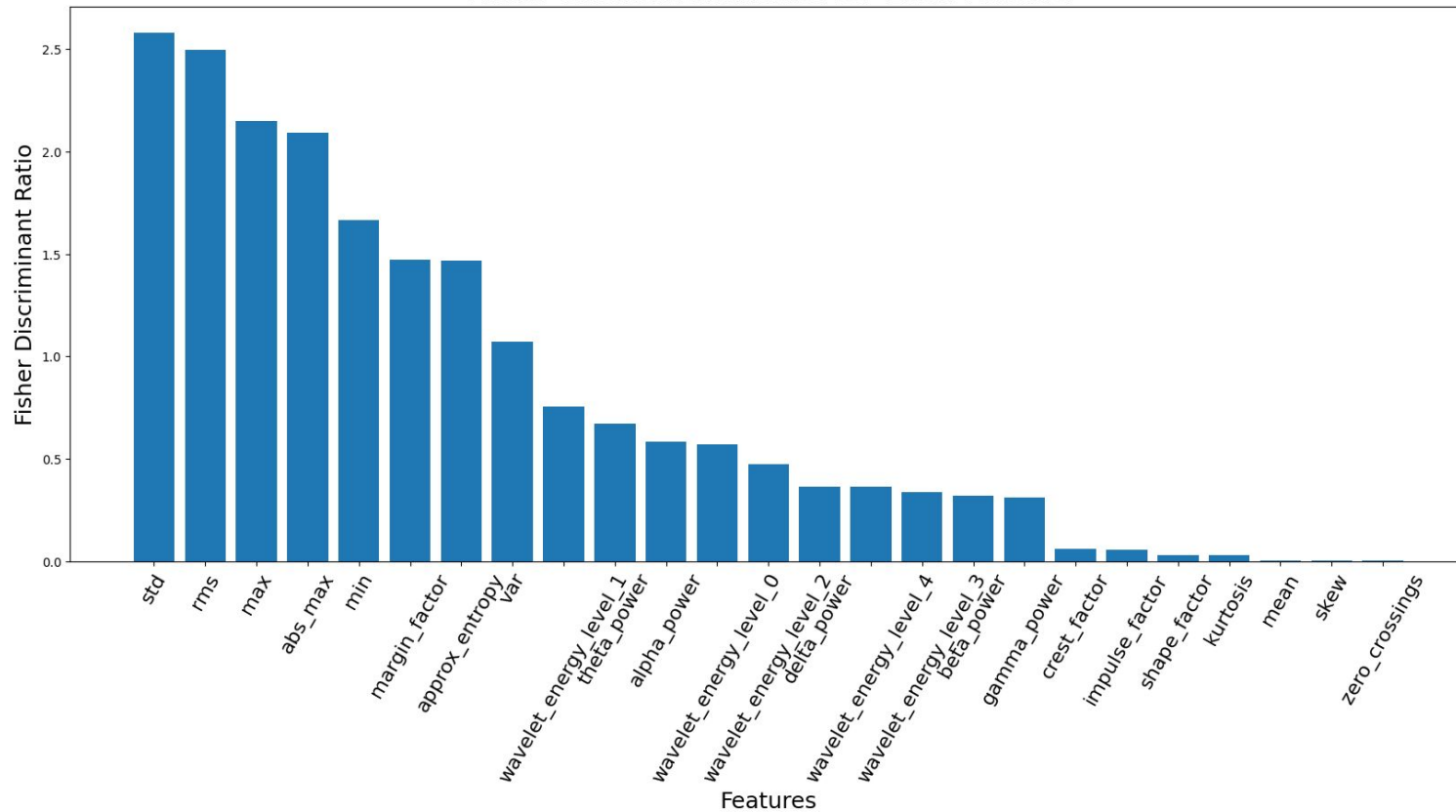
Capacidad Discriminante

- Características no normales
- Utilizamos razón discriminante de **Fisher**

$$FDR(X) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$



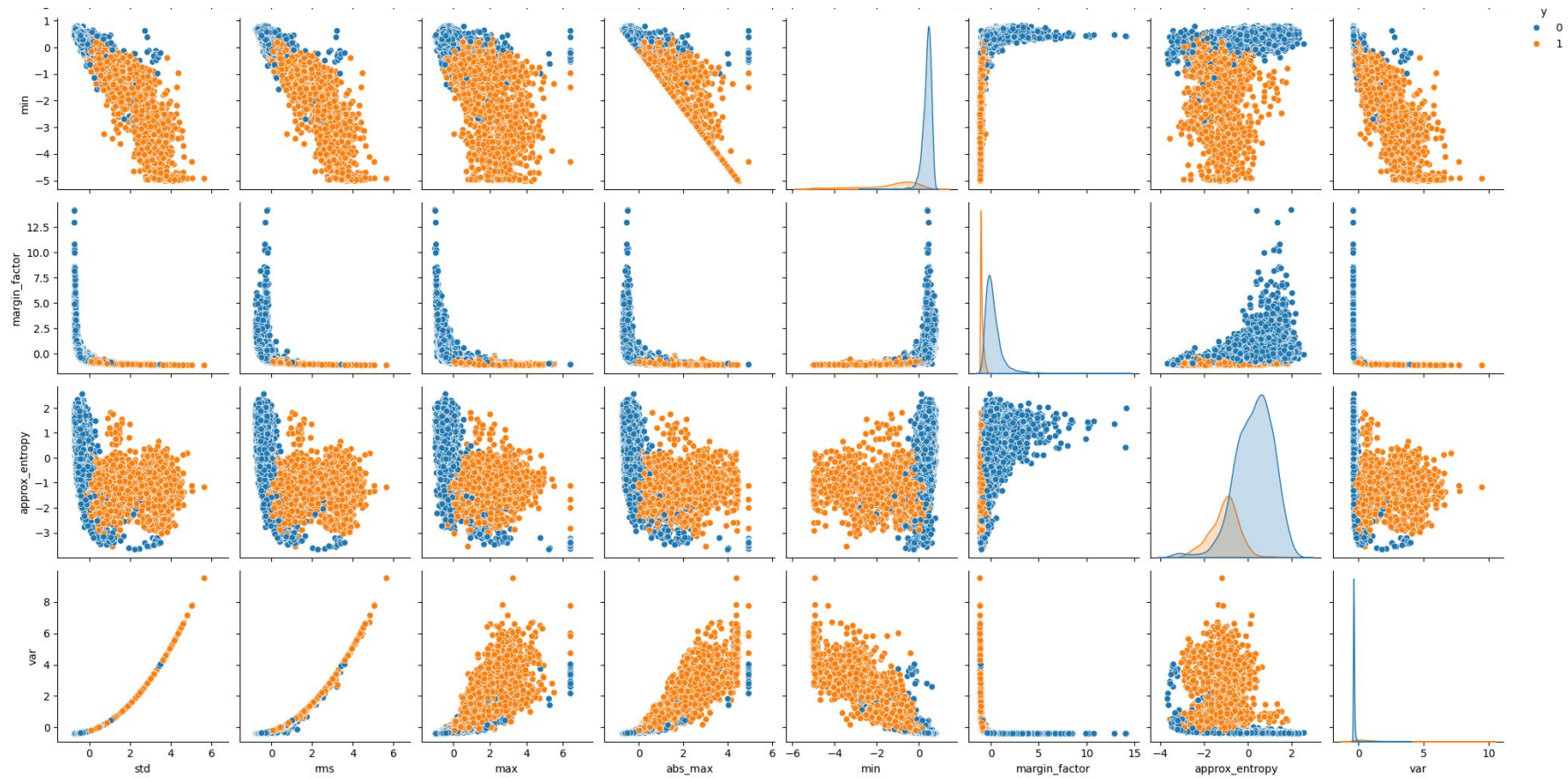
Fisher Discriminant Ratio for Each Feature



Selección de Características

Correlación Mutua

	std	var	min	max	rms	abs_max	crest_factor
std	1.000000	0.957911	-0.935807	0.920437	0.995684	0.964979	0.171243
var	0.957911	1.000000	-0.915336	0.853843	0.956332	0.923493	0.128806
min	-0.935807	-0.915336	1.000000	-0.788930	-0.935610	-0.940285	-0.227281
max	0.920437	0.853843	-0.788930	1.000000	0.912461	0.923646	0.287789
rms	0.995684	0.956332	-0.935610	0.912461	1.000000	0.965788	0.141358
abs_max	0.964979	0.923493	-0.940285	0.923646	0.965788	1.000000	0.322979
crest_factor	0.171243	0.128806	-0.227281	0.287789	0.141358	0.322979	1.000000
margin_factor	-0.523830	-0.388809	0.453055	-0.519505	-0.482824	-0.475422	-0.209298



Selección de Características

Criterio de selección basado en separabilidad y correlación

Criterio de separabilidad:

$$J_1 = \frac{\text{Tr}(S_b)}{\text{Tr}(S_w)}$$

S_w -> Varianza intra-clase

S_b -> Varianza inter-clase

Selección hacia adelante

1. Sea x_1 la característica con mayor $J_1(x)$.
2. Calcular **score** = $a1 \cdot J_1(x) - \max(a2 \cdot \text{corr}(x,z))$ para todo x no seleccionado y todo z en el subconjunto seleccionado.
3. Si la mejora es negligible o llegamos al máximo número de características.
 - a. Sí -> Fin
 - b. No -> Volver al paso 2

Ventaja: No necesitamos el modelo de clasificación a priori.

Desventaja: Dificultad en seleccionar correctamente valores de a.

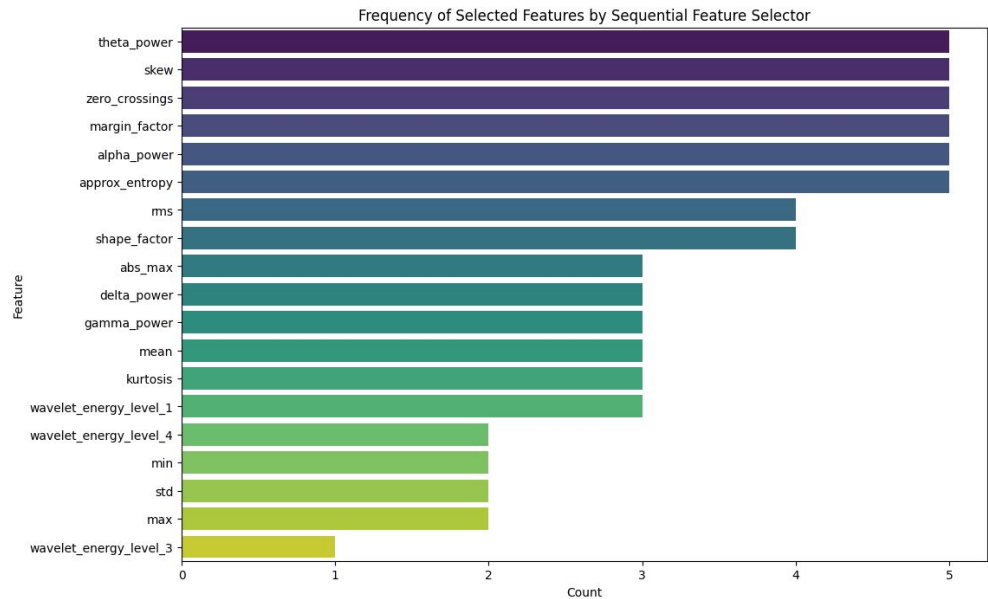
Mejores características según este criterio:

['std', 'rms', 'abs_max', 'max', 'min', 'approx_entropy', 'margin_factor', 'crest_factor', 'gamma_power', 'delta_power']

Selección de Características

Criterio de selección empírico

- Selección de características según si aumentan o empeoran las métricas de validación del modelo.
 - **Forward Selection:** Inicia con \emptyset y agrega características si mejoran el desempeño.
 - **Backward Elimination:** Inicia con todas las características y elimina aquellas que empeoran el desempeño.
- **Ventaja:** Se seleccionan las características que optimizan las métricas.
- **Desventaja:** Se introduce un sesgo inductivo al fijar el modelo a priori.



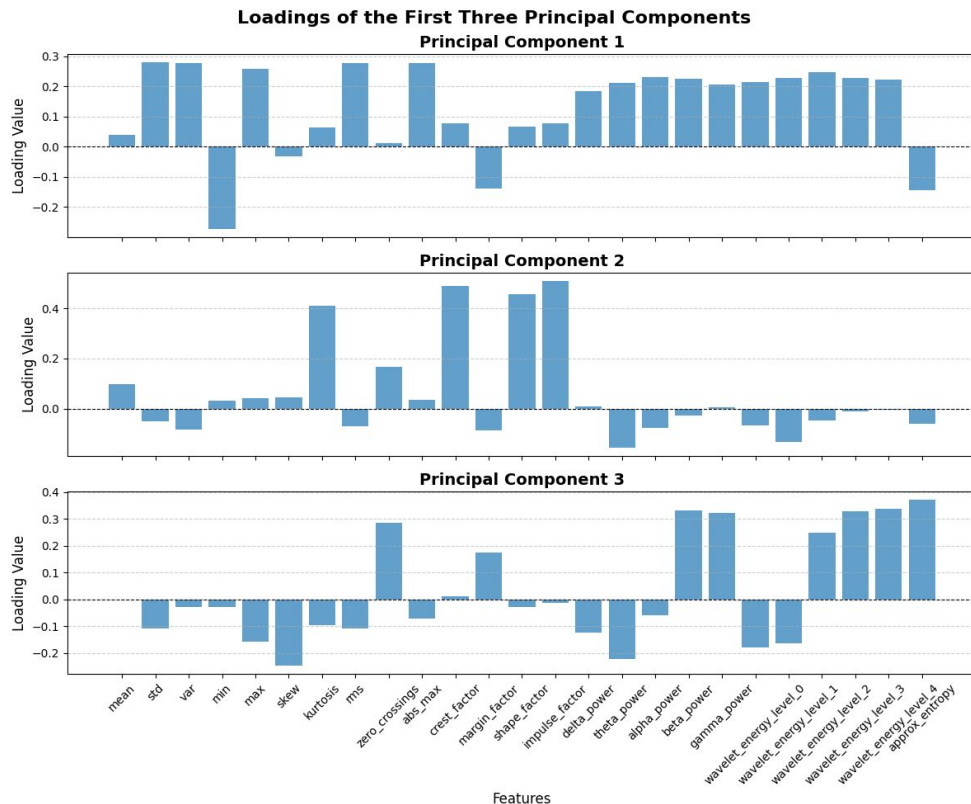
Separabilidad lineal en espacio de PC's

PC1 - Alta variabilidad positiva en todo el EEG.

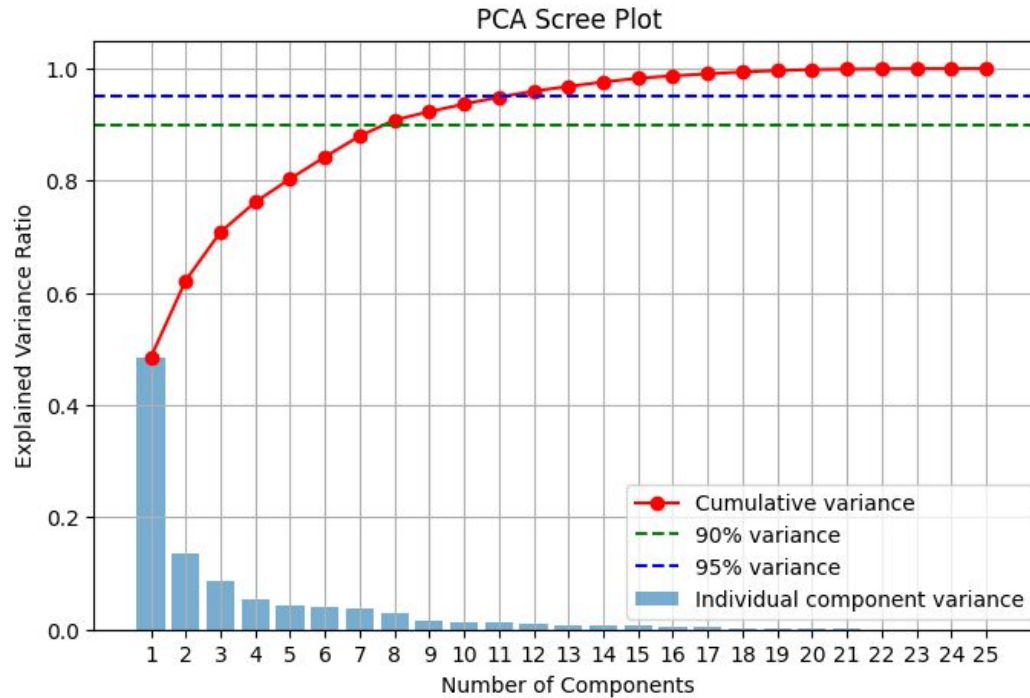
(HI zero_crossings, max, and all, rms, power bands. LOW min. and skew -> plano)

PC2 - Intensidad del pico más alto (HI kurtosis, shape, crest and margin factor)->cambios de estado cognitivo

PC3 - excitación destacando la actividad de alta frecuencia (gamma, beta) y la complejidad de la señal (approx entropy). LOW baja frecuencia (theta, delta) asociadas al descanso. Irregularidad propias de estados mentales activos.



Reducción de la dimensionalidad

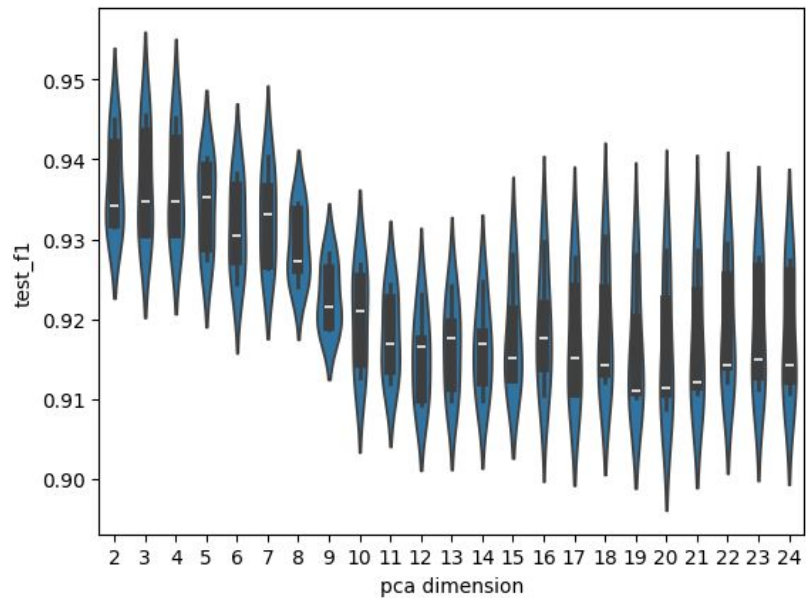




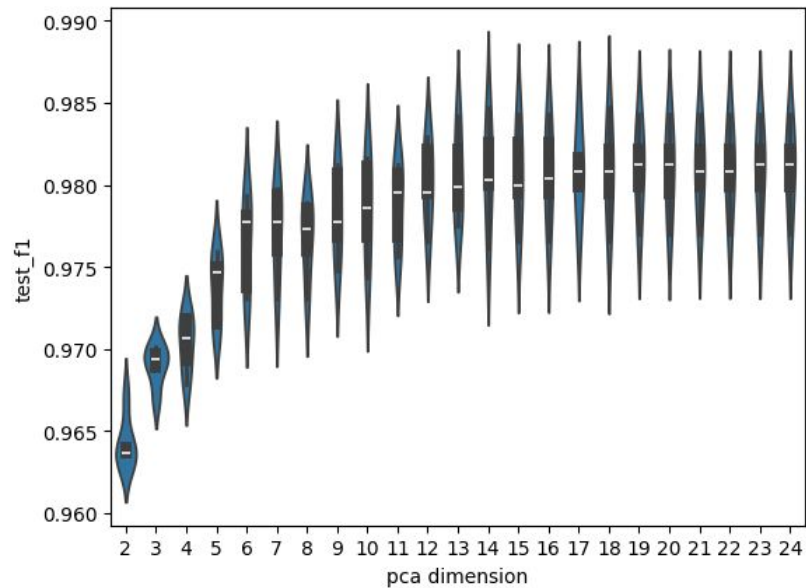
Evaluación

- Validación cruzada
- Métricas: F1-score, precisión, accuracy, recall
- Foco en la influencia de:
 - Modelo utilizado: Bayes o SVM.
 - Conjunto de características.
 - Dimensión de PCA.

Resultados

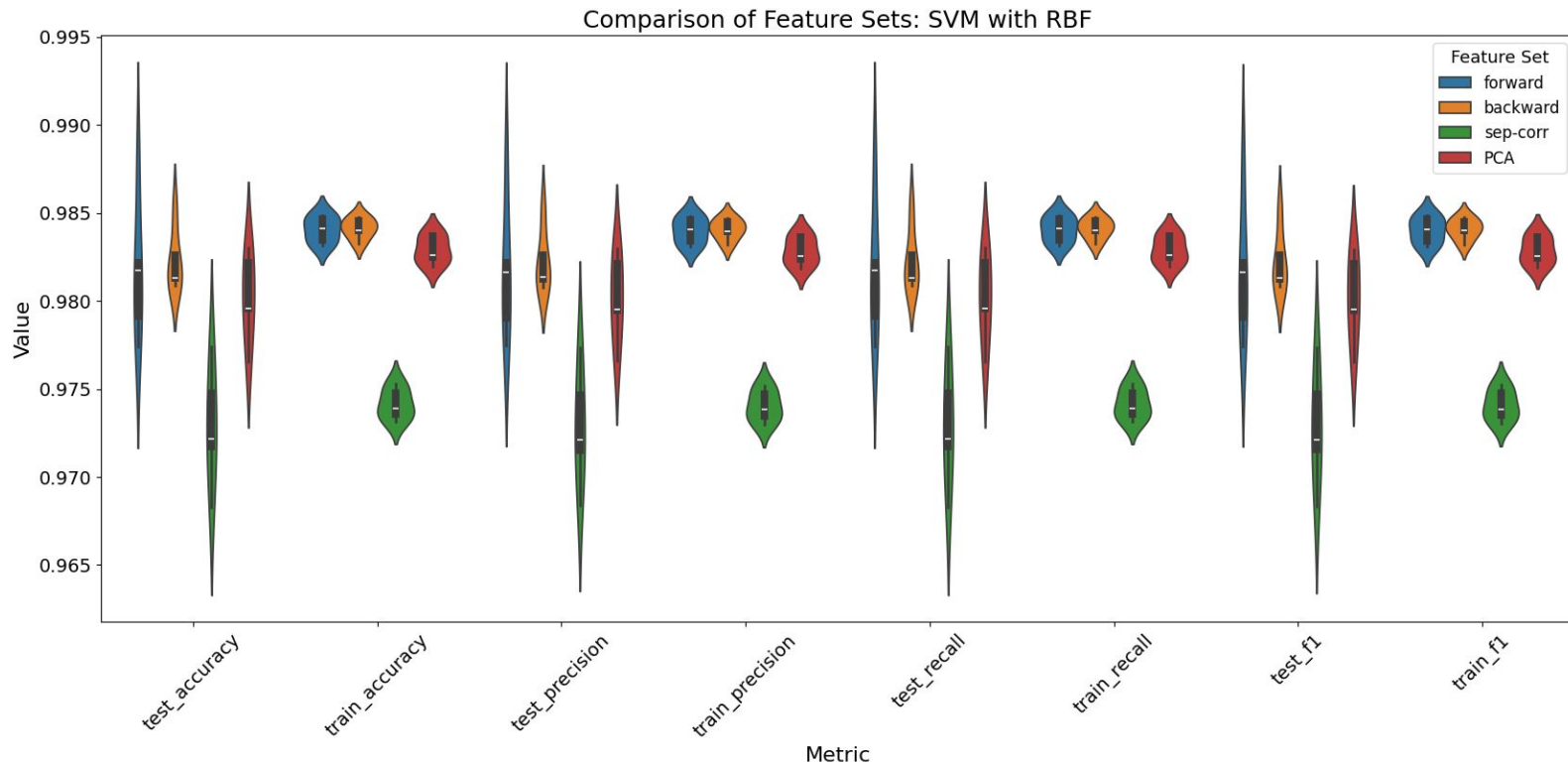


Clasificador Bayesiano

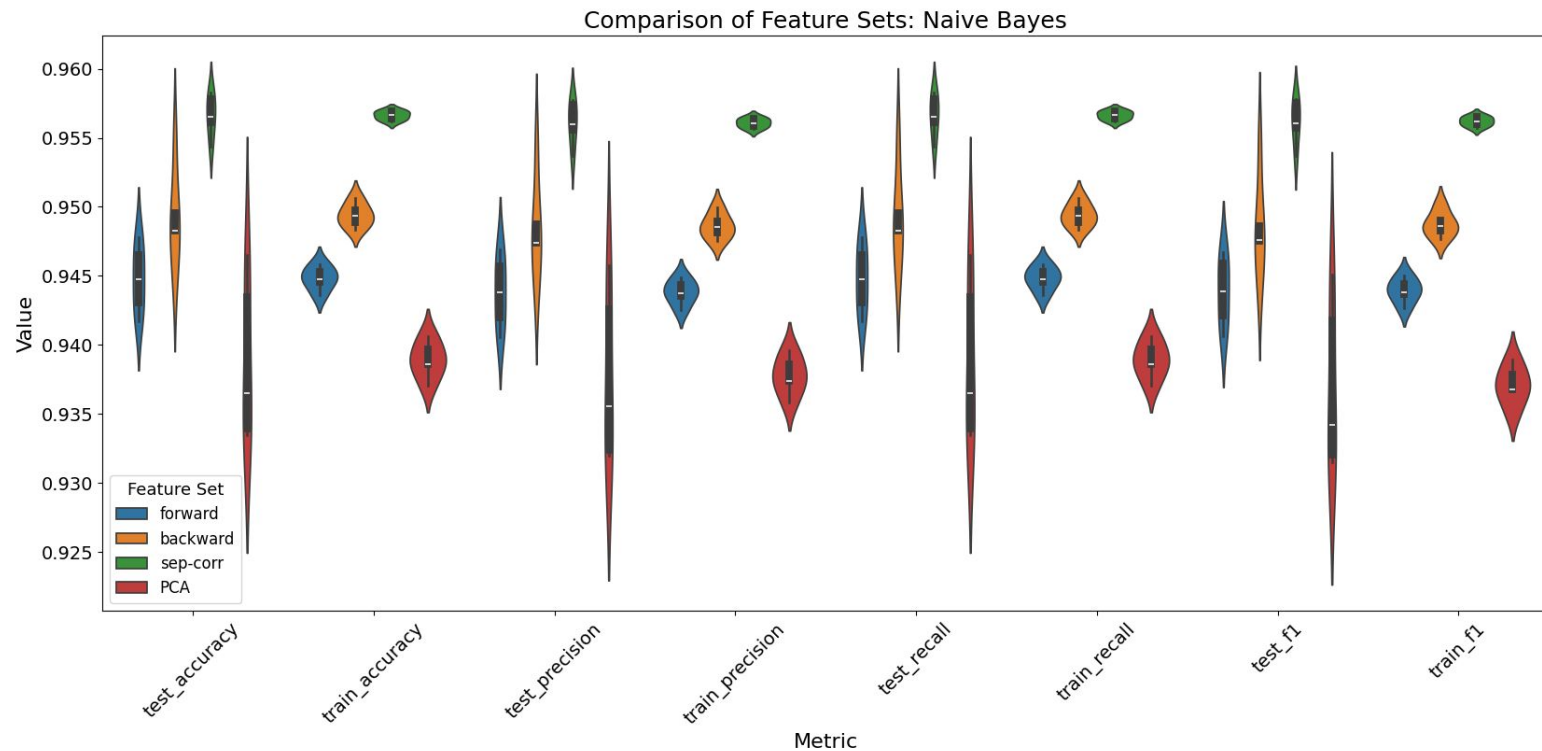


SVM con RBF kernel

Resultados



Resultados



Conclusiones y Trabajo Futuro

VENTAJAS:

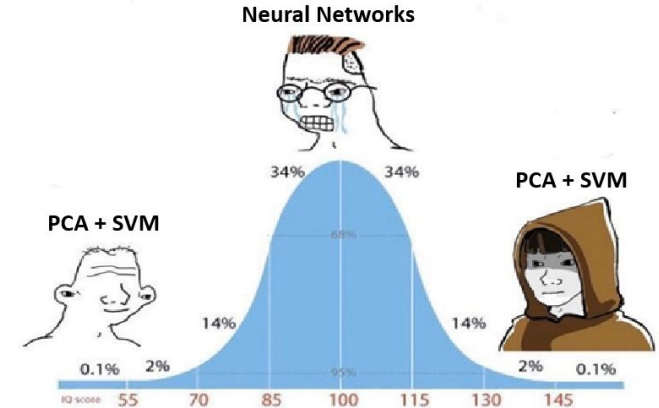
- Modelo simple + ALTA precisión en datos out-of-sample.
- Modelo generalizable para EEGs de mayor duración.
- Interpretabilidad -> ayudar al diagnóstico manual de epilepsia.

DESVENTAJAS:

- Pocos datos (500 Individuos) -> overfitting.
- A diferencia de modelos en academia, no diferencia entre tipos de epilepsia -> difícil de sistematizar tratamiento.
- Falsos Negativos.

MEJORAS:

- Patrones en errores
- Utilidad práctica: Bias para 0% Falsos negativos



Referencias



- [1] Delgado, K., & Ledesma, S. (2019). Análisis de electroencefalograma usando redes neuronales artificiales. *Acta Universitaria*, 29, 1-24.
- [2] Hamad, A., Hassanien, A. E., Fahmy, A. A., & Houssein, E. H. (2018). A hybrid automated detection of epileptic seizures in EEG based on wavelet and machine learning techniques. *arXiv preprint arXiv:1807.10723*.
- [3] Saby, J. N., & Marshall, P. J. (2012). The utility of EEG band power analysis in the study of infancy and early childhood. *Developmental neuropsychology*, 37(3), 253–273. <https://doi.org/10.1080/87565641.2011.614663>