

Departamento de Arquitectura y Tecnología de Computadores
Grado en inteligencia artificial

Análisis de Datos Biomédicos y Fisiológicos

Práctica sobre señales fisiológicas

Detección automática de activaciones del sistema nervioso simpático (SNS):

Deberás entregar un programa (jupyter notebook o similar) que detecte automáticamente las activaciones del sistema nervioso simpático construido a partir de los datos que descargarás de eGela (en el apartado datos_práctica) explicando y justificando todos los pasos seguidos como:

- Características extraídas para cada señal
- Tamaños de ventana utilizados
- Metodología de validación utilizada
- Métricas seleccionadas para la evaluación de los sistemas
- Métodos de aprendizaje y/o selección de características
- Resultados obtenidos y reflexión

La fecha final para **entregar** la práctica se **publicará en eGela**

Detección automática de la activación del sistema nervioso simpático

El objetivo final de esta práctica será entrenar modelos de aprendizaje automático capaces de detectar automáticamente la activación del sistema nervioso simpático (SNS). Es decir aglutinar los conocimientos adquiridos en la asignatura hasta el momento y juntarnos con los adquiridos en otras asignaturas del grado.

Para ello contamos con muestras adquiridas durante experimentos realizados en clase que hemos etiquetado debidamente. Contamos con **datos de 10 personas** con ECG y EDA grabado y anotado donde la activación del SNS está etiquetada con valor 1.

Los datos los podéis encontrar en eGela divididos en dos grupos (*Datos para la práctica. Primera parte* y *Datos para la práctica. Segunda parte*). Los nombres de los **ficheros** que contienen los datos están codificados según el siguiente esquema:

ESTRES001_20250312_001_xx.h5

ESTRES001_20250312_001_xx_etiquetas.csv

Donde el fichero con extensión h5 contiene las señales grabadas y la información sobre las condiciones en las que se han grabado, y el fichero con extensión csv contiene las etiquetas correspondientes a cada muestra. xx en cambio, corresponde a un identificador de persona. Dado el alto número de grabaciones mal etiquetadas entregadas, ha habido que utilizar muchos datos grabados con antelación en otras sesiones por lo que es muy probable que no encontréis los datos que habéis grabado en la muestra que se os proporciona.

Las señales que hemos adquirido son ECG (Electrocardiograma) y EDA (Actividad eletrodérmica). Ambas son señales o **series temporales que tendremos que adecuar** para alimentar los algoritmos de aprendizaje automático. Es decir partiendo de dichas señales tendremos que obtener muestras de tipo:

$X_{11}, X_{12}, X_{13}, X_{14} \dots X_{1n}, Y_1$

$X_{21}, X_{22}, X_{23}, X_{24} \dots X_{2n}, Y_2$

$X_{31}, X_{32}, X_{33}, X_{34} \dots X_{3n}, Y_3$

.....

$X_{m1}, X_{m2}, X_{m3}, X_{m4} \dots X_{mn}, Y_m$

Cada ejemplo lo obtendremos de una ventana de tiempo donde $X_{i1}, X_{i2}, X_{i3}, X_{i4} \dots X_{in}$ corresponderán a las características extraídas a partir de las señales ECG y EDA para la ventana i , e y_i corresponderá a la etiqueta de la misma ventana.

Los ejercicios realizados durante la asignatura, análisis y extracción de características de las señales EDA y ECG (y EMG) serán la base para el cálculo de las características.

Para la **preparación de las muestras** deberéis:

1- Seleccionar un **tamaño de ventana** y un **desplazamiento**. Por ejemplo ventanas de 5 segundos con desplazamiento de 1 segundo.

2- Para cada señal (EDA y ECG) y ventana, deberéis decidir las **características a extraer** que podrán ser:

- **Temporales**: medias, máximo, mínimo, rango, energía, área, pendientes...

- **Frecuenciales**: análisis para distintas bandas, espectro de frecuencias, energía, área ...

3- Además tendréis que decidir cuál es la **etiqueta** correspondiente a la ventana (0 o 1) en base al máximo entre las etiquetas de muestras correspondientes a dicha ventana.

Por ejemplo:

| | | | | |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 |
|---|---|---|---|---|

 --> 0

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 |
|---|---|---|---|---|

 --> 1

Será el mismo proceso de generación de muestras el que tendréis que aplicar a toda la base de datos tanto a las muestras de entrenamiento como a las de test.

Una vez codificada la base de datos, en el **proceso de aprendizaje** deberéis de utilizar:

- Para **entrenamiento** las **muestras** que obtuvieron durante las grabaciones **todos los grupos**. Es decir utilizaréis para entrenamiento los ejemplos extraídos de los ficheros ESTRES001_20240222_001_01.h5 . . . ESTRES001_20240222_001_08.h5 . En este proceso deberéis decidir en base a los conocimientos adquiridos en el grado:

- La estrategia de validación que consideréis más adecuada (cross-validation, hold-out...)

- La métrica de evaluación a utilizar (accuracy, AUC, F-measure...)

- El proceso de aprendizaje: preproceso (normalizaciones, selección de características..) y clasificadores a utilizar.

Con esto decidiréis cuál es la mejor configuración entre las que habéis probado: el mejor preproceso, las mejores características y el mejor clasificador.

- Para **test**: la víspera de la entrega se os facilitará una nueva muestra que servirá para **estimar la capacidad de generalización del sistema diseñado**. Para ello tendréis que construir el clasificador con mejores características utilizando la muestra de entrenamiento completa, y testearlo con los nuevos datos. El proceso de extracción de características para estos datos deberá de ser igual al utilizado con las muestras de entrenamiento. Los resultados obtenidos en este apartado permitirán comparar la bondad de los sistemas obtenidos por distintos grupos.

Entrega: se espera que entreguéis un Jupyter notebook o similar bien documentado que incluya además de código partes de documentación en las que se describen las decisiones tomadas o soluciones adoptadas y los resultados obtenidos (tanto los obtenidos con los datos de entrenamiento como los obtenidos con la muestra de test).