

PEC2: Debate y Actividad

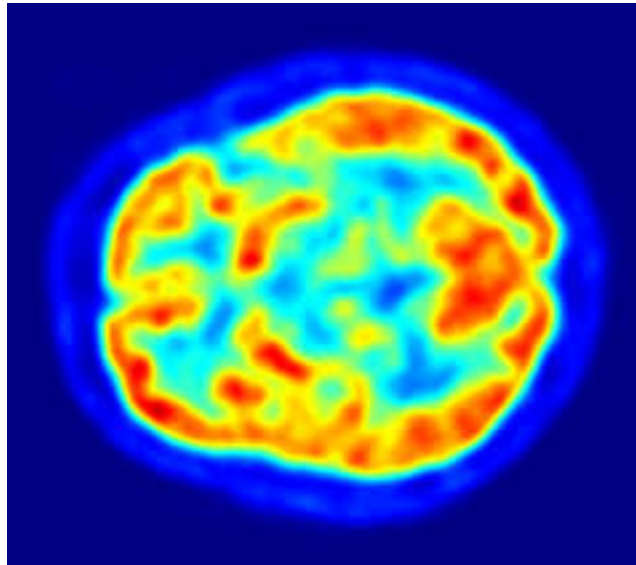
Machine Learning

La PEC2 consta de una **actividad** (70%) y de un **debate** (30%).

Actividad

Descripción

La enfermedad de Parkinson es un trastorno neurodegenerativo crónico y progresivo. Como las neuronas generadoras de dopamina en partes del cerebro se dañan o mueren, las personas comienzan a experimentar dificultades para hablar, escribir, caminar o completar otras tareas simples. Estos síntomas empeoran con el tiempo, lo que resulta en un aumento de su gravedad en los pacientes.



En esta actividad proponemos la implementación de una red neuronal profunda densa para la predicción de la severidad de la enfermedad de Parkinson a partir de datos captados por telemonitorización de la voz de los pacientes. Información complementaria sobre los datos se puede encontrar en: <https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring>.

Este conjunto de datos se compone de una variedad de mediciones biomédicas de voz de 42 personas con enfermedad de Parkinson en etapa temprana reclutadas para una prueba de seis meses de un dispositivo de telemonitorización para el monitoreo remoto de la progresión de los síntomas. Las grabaciones se capturaron automáticamente en los hogares de los pacientes.

Las columnas del data set contienen: ID del paciente, edad, sexo, intervalo de tiempo desde la fecha de reclutamiento inicial, motor_UPDRS, total_UPDRS y 16 medidas biomédicas de voz. Las variables que se utilizarán en la PEC son: total_UPDRS y 16 medidas biomédicas de voz. Cada fila corresponde a una de las 5.875 grabaciones de voz de estos individuos.

El objetivo principal de la actividad es predecir la severidad del Parkinson a partir de las 16 medidas de voz, donde la severidad del Parkinson se basa en una variable binaria que se define a partir de la binarización

de la variable `total_UPDRS`, indicando severidad cuando `total_UPDRS > 25` y no en caso contrario.

Enunciado

1. Cargar el fichero de datos `parkinsons_updrs.data`.
2. Realizar un descriptiva de las variables del dataset.
3. Normalizar mediante la transformación `min-max` las variables de las 16 medidas de voz.
4. Crear la variable binaria de severidad del Parkinson.
5. Separar train y test.
6. Implementar una red neuronal profunda densa para la predicción de la severidad. con dos capas ocultas de 10 y 10 nodos cada una. Adicionalmente, implementar una segunda red neuronal profunda densa con tres capas ocultas de 10, 20 y 10 nodos cada una.
7. Comparar el rendimiento de las dos redes neuronales en el conjunto test mediante sus curvas ROC y las metricas de calidad.

Informe de la actividad

Se entregaran dos ficheros:

1. Fichero ejecutable (.ipynb) que incluya un texto explicativo que detalle los pasos implementados en el script y el código de los análisis.
2. Informe (pdf) resultado de la ejecución del fichero (.ipynb) anterior.

Puntuacions de los apartados

Apartado 1 (5%), Apartado 2 (10%), Apartado 3 (5%), Apartado 4 (10%), Apartado 5 (5%), Apartado 6 (45%), Apartado 7 (20%)

Debate

El debate usa principalmente el artículo de revisión de Lecun et al. (ver abajo) como guía para tratar sobre Deep Learning. Para dar unas pautas en el debate se plantean las siguientes preguntas:

1. ¿Qué es el “Representation learning”?
2. ¿Cuál es la utilidad del “stochastic gradient learning”?
3. ¿Cuál es la función ReLU? y ¿Cuál es su utilidad?
4. En la práctica, ¿Son un problema los mínimos locales en el entrenamiento de grandes Neural Networks?
5. ¿Qué son las Convolutional Neural Networks? Descripción. Propósito y Aplicaciones.
6. ¿Qué son las Autoencoders Neural Networks? Descripción. Propósito y Aplicaciones.
7. ¿Qué son las Recurrent Neural Networks? Descripción. Propósito y Aplicaciones.
8. ¿Qué son las Memory Neural Networks? Descripción. Propósito y Aplicaciones.
9. Algoritmos generativos tales como Generative Adversarial Networks (GAN), Variational AutoEncoders (VAE)
10. Transformers.
11. Big data.
12. Métodos bioinspirados.
13. Otros temas en Inteligencia Artificial en Biociencias.

Las aportaciones se harán en el módulo de debates (NO es el foro) que tiene la UOC. Hay carpetas numeradas y separadas para cada pregunta. Además se ha añadido una última pregunta para comentar aspectos que no estén incluidos en las anteriores. Recordar que lo importante es hacer aportaciones efectivas, no de relleno. Es suficiente con hacer 5 contribuciones que tenga información/sugerencias/comparación/... RELEVANTE en el debate, incluyendo las referencias bibliográficas que son base o extensión de la contribución aunque siempre es mejor aportar el mayor número de contribuciones que se pueda.

Además del artículo de Lecun et al. se referencian dos artículos sobre Deep learning en bioinformática (ver abajo). Es material complementario para vuestra formación y puede dar origen a preguntas diferentes a las propuestas que podéis formular y contestar en la sección “Otros temas en Inteligencia Artificial en Biociencias.”

Referencias:

- Yann LeCun, Yoshua Bengio & Geoffrey Hinton (2015) Deep Learning. Nature Reviews, vol 521, 436-444
- Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. Molecular systems biology, 12(7), 878.
- Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. Briefings in bioinformatics, 18(5), 851-869.

Link de los artículos en la biblioteca de la UOC:

https://discovery.biblioteca.uoc.edu/discovery/fulldisplay?docid=cdi_proquest_miscellaneous_1684430894&context=PC&vid=34CSUC_UOC:VU1&lang=es&search_scope=MyInst_and_CI&adaptor=Primo%20Central&tab=Everything&query=any,contains,Deep%20learning&facet=rtype,include,articles&facet=searchcreationdate,include,2015%7C,%7C2015&facet=tlevel,include,peer_reviewed&offset=0

https://discovery.biblioteca.uoc.edu/discovery/fulldisplay?docid=cdi_pubmedcentral_primary_oai_pubmedcentral_nih_gov_4965871&context=PC&vid=34CSUC_UOC:VU1&lang=es&search_scope=MyInst_and_CI&adaptor=Primo%20Central&tab=Everything&query=any,contains,Deep%20learning%20for%20computational%20biology.%20Molecular%20systems%20biology

https://discovery.biblioteca.uoc.edu/discovery/fulldisplay?docid=cdi_proquest_miscellaneous_1826740389&context=PC&vid=34CSUC_UOC:VU1&lang=es&search_scope=MyInst_and_CI&adaptor=Primo%20Central&tab=Everything&query=any,contains,Deep%20learning%20in%20bioinformatics.%20Briefings%20in%20bioinformatics&offset=0

¿Qué hay que entregar?

Hay que subir a la actividad 2 del módulo de “Registro de EC” un fichero zip con:

1. Un documento pdf con las aportaciones realizadas en el debate indicando a que secciones corresponden.
2. Los dos ficheros correspondientes a la actividad.