



CURSO: CMP 5002 - DATAMINING
COLEGIO: POLITÉCNICO
Semestre: 1er Semestre 2022/2023

Proyecto 9: Ejercicio en equipos usando Deep Learning

Actividades

Del acto de investigación, implementación y experimentación:

1. Escoger una arquitectura *deep* por equipo (asignado aleatoriamente) y aplicarlo a la resolución de un problema de libre albedrío (buscar un dataset en el UCI repositories, Kaggle, etc) sin repetición entre equipos.
2. La arquitectura del modelo y el *dataset* debe ser único por equipo (investigar). **0.5 puntos**
3. Normalizar con *min-max* el *data set* escogido. **0.5 puntos**
4. Se debe usar la arquitectura seleccionada en conjunto con el método *stratified k-fold cross-validation* ($k=10$) para determinar el promedio y la desviación estándar de las siguientes métricas de evaluación de desempeño para problemas de clasificación: ACC, AUC, PRE, REC, F1 score, MCC (aplicar de proyectos anteriores), y *loss*. Para problemas de *forecasting*, el promedio y desviación estándar de: *Mean Square Error (MSE)*, *Mean Absolute Error (MAE)*, *R-Squared*, *Mean Absolute Percentage Error*, *Root Mean Squared Error*, *Normalized Root Mean Squared Error*, *Weighted Absolute Percentage Error*, *Weighted Mean Absolute Percentage Error*. **4 puntos**
5. Se deben utilizar elementos de optimización interna de la arquitectura seleccionada (evitar usar un modelo baseline de la internet, pues sería plagio). **1 punto**
6. Optimizar el modelo usando: *learning rate* = 10^{-4} con *ADAM* (o *SGD*, *RMSprop*, etc.) *optimizer*; *Epochs* = 1..200 con *checkpoint* cada 50 *epochs*. **0.5 puntos**
7. Los resultados deben ser mostrados en una tabla que contenga los modelos generados (y sus parámetros globales y particulares). Además, el desempeño de cada modelo por cada métrica evaluada. **1 punto**
8. Presentar la matriz de confusión resultante del mejor modelo seleccionado (maximizar AUC en *classification* o minimizar MSE en *forecasting*). **0.5 puntos**
9. Presentar los *plots* de AUC-ROC y PRE vs REC para el mejor modelo. **1 punto**
10. Presentar el *plot* de la función de *mean loss* vs *epochs* para el *training* y *test* (en el mismo *plot*). **1 punto**

Del acto de entrega:

- Se debe subir al D2L los códigos implementados y la presentación realizada en el tiempo de entrega.

Distribución:

- Deep neural networks (Fully connected) – (equipo2)
- Deep Convolutional neural networks (U-net, etc.) – (equipo1)
- Deep Recurrent Neural Networks (LSTM, GRU, RNN) – (equipo6)
- Deep Generative Adversarial Networks (GANs) - (equipo4)
- Deep Autoencoders (deep AE) – (equipo3)
- Vision Transformers (ViT) – (equipo5)

Nota: Errores de concepto en cualquier punto de control se penaliza con **2 puntos** más el valor del literal. La no entrega en tiempo y forma se penaliza con **0 puntos**. La no asistencia de algún integrante de un equipo a las presentaciones, se penaliza con **5 puntos** a todos los integrantes.

BUENA SUERTE.