

Aprendizaje de máquinas

Proyecto Parte 2

El proyecto del curso consiste en el desarrollo de diversos tópicos de Parte 2 del curso Aprendizaje de Máquinas. Se espera que vean tópicos vistos o por ver durante el curso. La tarea de cada proyecto depende del grupo y depende del tema elegido.

En particular, la tarea queda asignada por Proyecto (4) e Innovación (4). El Proyecto corresponde a la suma de letras de primer apellido de integrantes modulo 4 +1. Mientras que la Innovación corresponde a la suma de letras de segundo apellido de integrantes modulo 4 y luego sumado a 1. Por ejemplo si los integrantes tienen primer apellido Perez y McClane entonces seria $(5+7 \bmod 4)+1=3+1=4$; por lo tanto su tarea corresponde a Proyecto 4.

En el caso del Proyecto es incambiable. En el caso de innovación, pueden elegir otra pero deberá ser propuesta a cambio por correo antes de Entrega 1, y el profesor debe dar visto bueno. Cada innovación tiene diferente bonificación como premio a esfuerzo, indicada en Anexo; sin embargo, cada una es un proyecto válido en un curso como este.

En relación a entregas, se espera que cumplan múltiples hitos. En particular tendrá 2 entregas (en corchetes se anota el puntaje de cada ítem, que suman 5.5):

- a. Análisis y preprocesamiento de datos (Viernes 6/6 a las 11:59 pm, **peso: 30%**)
 - i. Definición de proyecto [0.5]
 - ii. Descripción de datos [1]
 - iii. Análisis de datos. *¿Cómo pueden comunicar las características principales de los datos?, por ejemplo distribución de clases así como estadísticas visuales como imágenes promedio por clase o imágenes muestra.* [1.5]
 - iv. Diseño de experimentos. Es decir que experimentos plantean realizar. Métricas, al menos dos. [1.5]
 - v. **Entrega de Lista de Integrantes en Excel (Grupos.xlsx). Por razones de operación deben entregar Lista indicando con X los integrantes de grupo.** [1]
- b. Implementación de algoritmos básicos (Viernes 27/6 a las 11:59 pm, **peso: 70%**)
 - i. Revisión de métodos relacionados con tarea, al menos 5 trabajos después del 2020 [0.5]
 - ii. Descripción de algoritmo(s) básico(s) para tarea. Debe ser al menos uno y que sea coherente con tarea [0.5]
 - iii. Descripción e implementación de algoritmo con innovación. [2]

- La idea de innovación es superar a modelo básico o experimentar con modelo nuevo en relación con problema.
- iv. Implementación de algoritmos en tarea [2.0]
 - v. Reporte de resultados obtenidos. Conclusiones. [0.5]

El documento en general tiene 0.5 puntos, considerando 1 punto base. Nótese que se ha indicado la fecha y ponderación de cada fase. La nota de este proyecto equivale a nota de proyecto de investigación (ver syllabus).

Cada innovación tiene una bonificación extra, pero ella solo se activará si logran presentar Entrega 2 con implementación funcional de algoritmo. El incumplimiento de innovación asignada, sin correo previo, implica una penalización de 2 puntos sobre nota final. El error de asignación de Proyecto o de Innovación se castiga con 1 punto sobre nota Final por cada uno.

Las entregas consisten en i. la documentación y ii. código de proyecto.

- Para documentación se sugiere usar Latex dado que es lenguaje de comunicación académico. Si lo hacen tendrán bonificación de 0.2 puntos solo en primera entrega.
- Para código se sugiere reportar usando notebook de Python. Para ello pueden usar Anaconda como Google colab.

En cada entrega deben entregar dentro de archivo ZIP:

1. La documentación, la cual contendrá al menos 7 páginas, preferentemente con impresiones de pantallas, y un máximo de 25 páginas por entrega. Se sugiere que entrega sea incremental al informe previo. Se recomienda el uso de Latex, pero no es obligatorio. Una página debe ser para caratula. El informe se debe entregar en formato fuente (Latex o Word) así como PDF.
2. Código fuente de avances de proyectos. **Nótese que es obligatorio; en otro caso tendrán nota mínima dado que es evidencia de trabajo realizado.**

Se debe enviar archivo ZIP a buzón de curso. Opcionalmente pueden usar servicios en la nube si los archivos fueran muy voluminosos. El mínimo/máximo de alumnos por grupo es de 2/5 integrantes. La copia es penalizada siguiendo indicatrices de UNAB. El uso de IA generativa debe ser hecha con precaución y seriedad, toda salida debe ser corroborada. Cualquier consulta práctica se hará con el profesor de curso ya sea vía mail o en oficina.

Saludos y suerte en este reto,

Billy

ANEXO:

A). Temas propuestos

Este año la propuesta va por temas médicos, y para ello se usará las bases de datos estandarizadas MedMNIST. Los datos deben ser al menos de resolución 64*64 y se deben descargar de <https://medmnist.com/>.

Código inicial -para descarga de datos- operativo a la fecha:

https://github.com/MedMNIST/MedMNIST/blob/main/on_medmnist_plus.md

El modelo base y propuesto debe ser diferente al que aparece en código inicial mostrado encima naturalmente (o mejor dicho no considerar MLP, salvo que implementen además otro modelo).

Los proyectos a elegir son:

P1. Reconocimiento visual de enfermedad respiratoria: ChestMNIST

¿Dada una radiografía de tórax, cuál es la enfermedad respiratoria subyacente?

P2. Detección temprana de cáncer de mama: BreastMNIST

¿Se puede detectar la presencia de cáncer en una imagen de ecografía mamaria?

P3. Diagnóstico de anomalías pulmonares pediátricas: PneumoniaMNIST

¿Esta radiografía infantil muestra signos de neumonía?

P4. Clasificación de imágenes de piel: DermaMNIST

¿Qué tipo de lesión dermatológica muestra esta imagen?

- Pueden también realizar propuestos por ustedes previa aprobación vía correo del Profesor. En especial si es data inédita.

B). Innovación

I1. Uso de ensamble como stacking u otro -diferente a MoE- [extra: 0.2 puntos en entrega 2]

I2. Uso de clustering, y luego aplicar red en cada cluster. [extra: 0.3 puntos en entrega 2]

I3. Uso de Mezcla de Expertos [extra: 0.4 puntos en entrega 2]

14. Considerar como salida una instancia de cada clase –"medoide" por ejemplo-, o vector diferente a clase (que es un entero). Clasificación final usando Distancia de Mahalanobis, por ejemplo. [extra: 0.5 puntos en entrega 2]

*Pueden probar otra innovación. Si es de un modelo ya implementado en librerías (por ejemplo Visual Transformers), no habrá bonificación dada mucho mayor simplicidad. En caso de otra innovación de acuerdo a juicio de profesor puede haber bonificación con máximo de 0.2 puntos y se debe indicar en Entrega 1.