

TD VI: Inteligencia Artificial

Trabajo práctico 3 (2025 2^{do} semestre)

Clasificación de audio

El propósito de este trabajo práctico es evaluar y analizar diversos hiperparámetros durante el proceso de entrenamiento de una red neuronal. Se buscará comprender cómo estos hiperparámetros afectan el rendimiento del modelo, permitiendo así la identificación y selección del conjunto óptimo de parámetros. A través de esta evaluación, se fomentará una comprensión más profunda de la influencia de los hiperparámetros en el aprendizaje automático y su impacto en la eficacia del modelo.

Base de datos:

El trabajo práctico se construye sobre la base de datos Speech Commands ¹ que puede ser descargada directamente desde pytorch y explorado kaggle <https://www.kaggle.com/datasets/yashdogra/speech-commands>

El mismo contiene audios de 1sec cada un y sus clasificaciones como "yes", "no", "up", "down", "left", "right", "on", "off", "stop", and "go".

Notebook ejemplo:

Desde el campus se podrá descargar una notebook “Implementacion_base_TP3.ipynb” la cual le servirá como ejemplo de lectura y reproducción de audios. En la notebook encontrará un clasificador de los audios basado en un MLP simple, cuyo funcionamiento no es óptimo.

Se pide:

- 1) **Configuración (1 punto):** Crearse una cuenta en Weights and Biases (o en su *board* favorito) y *linkear* la notebook a este. Cada experimento deberá contener nombres dientes y almacenar los (hiper)parámetros de configuración del mismo. Deberá separar los sets de datos en entrenamiento, validación y test. Utilizando solamente los sets de entrenamiento y validación, registrar la *loss* en *train* y validación en cada iteración que considere conveniente.
- 2) **Arquitectura (1 pts):** Realizar experimentos variando cantidad de capas densas, nodos, *hidden layers* y reportar el mejor y peor experimento. ¿Qué estrategia utilizaron? ¿Qué resultados obtuvieron?
- 3) **MEL vs WAV (1 pts):** Evaluar el funcionamiento comparando ambos formatos. ¿Cómo podemos ingresarlos tanto en un MLP como en una CNN?
- 4) **Funciones de activación (1 punto):** Realizar experimentos variando distintas funciones de activación y reportar el mejor y peor experimento. Explicar.
- 5) **Optimizadores (1.5 pts):** Realizar experimentos evaluando distintos optimizadores, *schedulers* y reportar el mejor y peor experimento. ¿Qué estrategia utilizaron? ¿Qué resultados obtuvieron?
- 6) **Arquitectura CNN (1.5 punto):** Extienda el análisis utilizando capas convolucionales y reportar el mejor y peor experimento. ¿Cómo se compara?

- 7) **Entrenamiento (1 punto):** Realizar experimentos evaluando distintos *batch-sizes*, epochs y reportar el mejor y peor experimento. ¿Qué resultados obtuvieron?
- 8) **Regularización (1 punto):** Evaluar alguna(s) técnica(s) de regularización. ¿Qué resultados obtuvieron?
- 9) **Input (1 punto):** Comparar el resultado de utilizar espectrogramas contra utilizar waveforms.
- 10) **Evaluación final (1 punto):** Evaluar el mejor modelo, es decir, el mejor de los ejercicios anteriores:
 - a) con el set de test.
 - b) con audio grabando sus voces.

Entregables:

- 1- Se debe entregar un informe explicando cada uno de los ejercicios, con los correspondientes gráficos obtenidos en Weights and Biases. Los gráficos deberán mostrar el gráfico con la loss y la *accuracy*. Para cada inciso, muestre un gráfico de barras donde se vea el mejor y el peor de los experimentos.
- 2- Metadata de los experimentos mencionados consistiendo en todos los hiper-parámetros evaluados y un nombre descriptivo de la arquitectura utilizada.
- 3- Notebook que evalúe y confirme el reporte del modelo ganador. El mismo debe ejecutarse de punta a punta sin errores y debe ser claro y legible para una persona ajena al grupo.

Aclaraciones:

- 1- Puede modificar el orden de ejecución de los ejercicios y utilizar la mejor configuración de un inciso para los incisos siguientes.
- 2- Si bien en cada inciso es posible modificar las demás variables, es importante que cuando reporte el mejor y peor experimento ambos se encuentren bajo las mismas condiciones. En caso de requerir un setting distinto para la evaluación, justificar.

Entrega:

El trabajo práctico se deberá resolver de a grupos de **3 integrantes**.

Fechas y modalidad de entrega

- El **informe** y el **código** podrán entregarlos hasta el domingo **9 de noviembre** a las 23:59:00.
- El **informe** debe ser entregado en formato **PDF**.
- Los **metadata** deberán ser archivos jsons (uno por experimento).
- Los archivos deben ponerse dentro de una carpeta llamada **tp3-gxx**, donde **xx** sea reemplazado por el número de grupo; por ejemplo, **tp3-g01**. La versión ZIP de esta carpeta



debe subirse a la tarea llamada **TP3 | Entrega** en la página de la materia en el Campus Virtual.

- **Sólo 1** integrante del grupo debe realizar la **entrega**.
- Para realizar sus consultas sobre este TP, cuentan con el foro llamado **TP3 | Consultas** en la página de la materia en el Campus Virtual. Todas las dudas que surjan en relación al TP3 envíenlas exclusivamente a este foro; no usen ningún otro. Esto debe ser así porque este es el foro que está configurado de forma que los mensajes enviados lleguen únicamente al cuerpo docente y a sus compañeros de grupo. En otras palabras, si un integrante de un grupo envía una pregunta por acá, tanto esa pregunta como la respuesta, luego dada por el cuerpo docente, podrán ser vistas sólo por los integrantes del grupo en cuestión.