



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
IIC-3697 APRENDIZAJE PROFUNDO

PROYECTO DE CURSO: ANÁLISIS DE TEXTO CON APRENDIZAJE PROFUNDO

Fecha de Entrega: Martes 20 de Julio

Estimad@s, llegó la hora de comenzar el proyecto de curso, en el cual podrán poner en práctica los conocimientos que han adquirido este semestre sobre aprendizaje profundo o deep learning (DL). Dado que Aprendizaje Profundo es un curso nivel 3000, es decir, nivel postgrado, también en este proyecto deberán ejercitar su capacidad para realizar estudio independiente e investigación en esta área.

En esta oportunidad podrán desarrollar su proyecto de curso en grupos de 2 o 3 integrantes. Tienen libertad para formar los grupos, si alguien necesita ayuda, por favor, enviar email al profesor.

Este semestre el proyecto estará centrado en desarrollar un modelo de aprendizaje profundo para el set de datos SQuAD 2.0 (Stanford Question Answering Dataset, <https://rajpurkar.github.io/SQuAD-explorer/>). Este set de datos consiste de más de 100.000 pares pregunta-respuesta contruidos en forma manual en base a documentos de Wikipedia. Adicionalmente, contiene 50.000 distractores, los cuales consisten en preguntas para las cuales no hay respuesta que se desprenda del documentao asociado, por tanto, el sistema debe abstenerse de dar una respuesta para estas preguntas.

1 Nivel de Complejidad Esperado

El proyecto es una parte importante del curso, tanto en términos de su metodología como de su sistema de evaluación. IIC-3697 es un curso de 10 créditos semanales, de acá en adelante parte sustancial de este tiempo debe ser dedicado al desarrollo de un buen proyecto de curso. A grandes rasgos es posible identificar 4 posibles niveles de complejidad para el proyecto. En orden creciente de complejidad:

- Implementar y evaluar algún método existente, ya sea algún algoritmo visto en clases o alguno que ustedes hayan encontrado en su revisión bibliográfica. En su reporte debe incluir un análisis de rendimiento bajo distintas condiciones, descripción de ventajas y limitaciones, y una comparación respecto a modelos alternativos.
- Proponer, implementar y evaluar alguna mejora a algún método existente. En este caso, la comparación de rendimiento sería con respecto al algoritmo base al cual se agregó la mejora.
- Combinar varios métodos con objeto de crear un sistema más robusto (ensamble).
- Realizar una innovación relevante a un método existente o desarrollar un nuevo modelo.

Lograr un buen nivel en cualquiera de los puntos anteriores garantizará una buena calificación. En todos los casos, especialmente en los puntos a) y b), será fundamental respaldar el estudio con un adecuado análisis de los resultados obtenidos, en términos de análisis de sensibilidad y comparación con métodos alternativos. En este sentido, limitar el proyecto de curso a bajar una herramienta de la web y probar su rendimiento no es suficiente para obtener una buena calificación en el proyecto de curso.

Importante: si bien el proyecto de curso se establece en modo desafío (challenge), es decir, se trabaja sobre un mismo problema, con lo cual los rendimientos pueden ser rankeados según la exactitud alcanzada, un trabajo con un nivel de rendimiento más bajo de lo esperado, pero que explore en forma metódica y crítica

una solución interesante, también tendrá una buena nota final. En este sentido, tengan presente que el obtener un buen resultado cuantitativo es una característica deseada pero no nuestro objetivo central, el cual es que desarrollen un proyecto de curso que aporte en su proceso de formación en las temáticas del curso. Por el contrario, un trabajo que obtenga un excelente resultado cuantitativo, pero consista en simplemente copiar una implementación pública, será penalizado con una nota deficiente.

2 Entregas

El proyecto incorpora 2 entregas.

Informe de Avance. Fecha entrega: martes 29/06

Este informe de avance no tiene nota pero será tomado en cuenta al momento de evaluar el informe final. El objetivo de este informe es demostrar que el desarrollo del proyecto se ha realizado en forma sistemática durante el período de tiempo asignado. El formato de entrega es un breve documento en pdf indicando los avances a la fecha y un archivo con los avances del código implementado, no importa que este último no sea aún funcional.

Entrega Final. Fecha entrega: martes 20/07

El informe final debe ser documentado en formato paper. En la página web del curso se publicarán templates en latex y MS-Word que pueden ser usados para este fin. El documento final debe incluir las siguientes secciones:

- Breve introducción describiendo las aristas principales de su propuesta.
- Revisión bibliográfica que incluya al menos 2 trabajos relevantes.
- Descripción del modelo propuesto, incluyendo un diagrama de bloques y una breve descripción de cada bloque.
- Resultados experimentales.
- Descripción de las principales conclusiones de su trabajo.
- Detalle de las referencias en formato paper.

Adicionalmente, deben entregar un python notebook con su código y los archivos de su modelo.