Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería de Transporte y Logística

ICT3115 – Sistemas Urbanos Inteligentes (I/2021)

Tarea 2: *Transfer Learning* y *Catastrophic Forgetting*

Introducción

En esta tarea tendrá la oportunidad de experimentar con el uso de redes neuronales convolucionales para el procesamiento de datos visuales. En particular, deberá entrenar CNNs para clasificación de imágenes sobre múltiples conjuntos de datos y evaluar cuanto de los aprendido en ellos es transferido u olvidado. Para el desarrollo se utilizará el framework Pytorch sobre la plataforma Colab de Google. Dado que existe abundante material disponible en línea para colaborar con el desarrollo de la tarea, se espera que todo recurso externo utilizado, sea este código, librerías o papers, esté debidamente indicado.

Set de datos

Las fuentes primarias de datos para entrenar los modelos serán los conjuntos MNIST, Fashion MNIST, CIFAR10 y SVHN. Cada uno de estos conjuntos contiene imágenes de distintos elementos, ya sea en escala de grises o color, y una cantidad predefinida de categorías a utilizar para la clasificación. Todos estos conjuntos se encuentran disponibles directamente en Pytorch.

Para crear sets de entrenamiento, validación y test independientes, utilice las funcionalidades de `scikit-learn`, como han sido indicadas en los ejemplos de código. Si el conjunto de datos ya tiene definidos estos conjuntos, respételos.

Modelos

Para esta tarea, debe utilizar modelos de redes convolucionales profundas como los descritos en el capítulo 3 del curso, sin limitante en relación al tipo de capas que se puede utilizar (convolucional, densa, *dropout*, *batchnorm*, etc). Se recomienda revisar bibliografía relacionada para esto. Considere además

preprocesar las entradas de acuerdo a lo utilizado en la literatura (por ej., restar la media, normalizar en el rango [0,1], etc.). No hay problema en basarse completamente en modelos propuestos previamente en la literatura o en tutoriales. En cualquier caso, debe justificar sus elecciones.

Tareas a realizar

Para la tarea se espera se espera que realice al menos las siguientes actividades:

- Entrenamiento y transferencia: entrene para cada set de datos el mejor modelo posible, y luego realice un proceso de *transfer learning* para todas las combinaciones, es decir, para cada modelo entrenado en un set de datos, debe realizar la transferencia a los restantes cuatro conjuntos de datos. Para la transferencia, considere la sustitución y posterior reentrenamiento de la capa de clasificación, deje todo el resto de la red sin modificar. Para cada set, compare el rendimiento de los modelos utilizando *accuracy*. Analice los resultados y discuta sobre la transferibilidad relativa de cada conjunto de datos.

- Entrenamiento secuencial con *fine-tuning*: elija una secuencia arbitraria de los sets de datos y entrene un modelo de manera secuencial en ellos, es decir, una vez entrenado el modelo en el set de datos t, respalde y sustituya la capa de clasificación, y utilice los pesos del resto de las capas como punto de inicio en un nuevo entrenamiento en el set $t+1$. Compare los rendimientos con los obtenidos en el ítem anterior y discuta sobre la transferibilidad relativa de cada conjunto de datos.

- *Catastrophic forgetting*: En base a los modelos del ítem anterior, evalúe el nivel de olvido inducido en los pesos al realizar el proceso de entrenamiento secuencial con *fine-tuning*. Para esto, considere la métrica de rendimiento Acc_T, t, que mide el *accuracy* en el set de datos t, luego de ser entrenado en el set T. En palabras simple, después de entrenar el modelo secuencialmente hasta llegar al set T, se sustituye la capa de clasificación de este por la obtenida para el set t y se evalúa el rendimiento en este set. Dada esta definición, reporte el nivel de olvido en base a las métricas *Mean Accuracy* (MAcc) y *Backward Transfer* (BWT):

$$MAcc = \frac{1}{T}\sum_{t=1}^{T} Acc_{T,t}$$

$$BWT = \frac{1}{T-1}\sum_{t=1}^{T-1} Acc_{T,t} - Acc_{t,t}$$

  Analice y discuta los resultados en base a las características de los sets de datos.

- Revise la literatura sobre olvido catastrófico e implemente algún mecanismo para controlarlo. Compare los nuevos resultados con los obtenidos en el ítem anterior.

Desarrollo y entrega

La tarea puede desarrollarse de manera individual o en parejas, utilizando el framework Pytorch para Python. Se recomienda utilizar la plataforma Google Colab con el fin de facilitar la instalación de librerías. Esta plataforma permite utilizar gratuitamente una GPU para el entrenamiento por intervalos de 12 horas continuos. En el *notebook* desarrollado debe ir tanto el código como un informe (preferiblemente intercalados), donde se expliquen los pasos realizados, se analicen los resultados y se planteen conclusiones. La entrega de la tarea tiene como fecha límite el mertes 08 de junio a las 23:59, a través del buzón que se habilitará en el sitio del curso. Para fines de corrección, se revisará la última versión entregada.

Política de Integridad Académica

Los alumnos de la Escuela de Ingeniería deben mantener un comportamiento acorde al Código de Honor de la Universidad:

"Como miembro de la comunidad de la Pontificia Universidad Católica de Chile me comprometo a respetar los principios y normativas que la rigen. Asimismo, prometo actuar con rectitud y honestidad en las relaciones con los demás integrantes de la comunidad y en la realización de todo trabajo, particularmente en aquellas actividades vinculadas a la docencia, el aprendizaje y la creación, difusión y transferencia del conocimiento. Además, velaré por la integridad de las personas y cuidaré los bienes de la Universidad."

En particular, se espera que mantengan altos estándares de honestidad académica. Cualquier acto deshonesto o fraude académico está prohibido; los alumnos que incurran en este tipo de acciones se exponen a un procedimiento sumario. Ejemplos de actos deshonestos son la copia, el uso de material o equipos no permitidos en las evaluaciones, el plagio, o la falsificación de identidad, entre otros. Específicamente, para los cursos del Departamento de Ciencia de la Computación, rige obligatoriamente la siguiente política de integridad académica en relación a copia y plagio: Todo trabajo presentado por un alumno (grupo) para los efectos de la evaluación de un curso debe ser hecho individualmente por el alumno (grupo), sin apoyo en material de terceros. Si un alumno (grupo) copia un trabajo, se le calificará con nota 1.0 en dicha evaluación y dependiendo de la gravedad de sus acciones podrá tener un 1.0 en todo ese ítem de evaluaciones o un 1.1 en el curso. Además, los antecedentes serán enviados a la Dirección de Docencia de la Escuela de Ingeniería para evaluar posteriores sanciones en conjunto con la Universidad, las que pueden incluir un procedimiento sumario. Por "copia" o "plagio" se entiende incluir en el trabajo presentado como propio, partes desarrolladas por otra persona. Está permitido usar material disponible públicamente, por ejemplo, libros o contenidos tomados de Internet, siempre y cuando se incluya la cita correspondiente.