

Trabajo práctico: Optimización

Ph. D. Saúl Calderón Ramírez
Instituto Tecnológico de Costa Rica,
Escuela de Ingeniería en Computación,
PAttern Recognition and MACHine Learning Group (PARMA-Group)

4 de octubre de 2024

El presente proyecto trata sobre la implementación en pytorch de distintos algoritmos de optimización.

- **Fecha de entrega:** Lunes 21 de Octubre
- **Modo de trabajo:** Grupo de tres/dos personas.
- **Tipo de entrega:** digital, por medio de la plataforma TEC-digital.

Para la documentación externa de este proyecto, incluya la explicación de cada método, y al menos 2 pruebas unitarias debidamente documentadas por cada uno. Para la documentación interna utilice un estándar de Doxygen <https://tinyurl.com/55hxcd7r>. La documentación externa debe realizarse en un documento pdf generado por latex, y la implementación debe entregarse en un notebook de jupyter.

1. Optimización de funciones

Para las siguientes funciones:

$$f_0(x, y) = x^2 + y^2$$

$$f_1(x, y) = (1,5 - x + xy)^2 + (2,25 - x + xy^2)^2 + (2,625 - x + xy^3)^2$$

$$f_2(x, y) = 0,26(x^2 + y^2) - 0,48xy$$

con $x_1, x_2 \in [-10, 10]$.

Para todos los algoritmos, ejecutelos por $P = 25$ iteraciones e inicialice las soluciones en el rango $x_1, x_2 \in [-10, 10]$.

1. **(20 puntos)** Según tales gráficas, grafique las funciones usando la función *meshgrid* y *contour*, y distinga si las funciones son convexas o no, y los puntos mínimos y regiones o puntos silla.
2. **(40 puntos)** Implemente el algoritmo del **RMS prop**. para encontrar el punto mínimo en pytorch, de la forma mas vectorial posible. Implemente las siguientes pruebas.
 - a) Para cada funcion, calibre los hiper-parámetros del algoritmo, mostrando el proceso de calibrado usando las gráficas de aprendizaje , y reporte los mejores valores encontrados.
 - 1) Realice el proceso de calibración usando un framework como *optuna* <https://optuna.org/> o *weights and biases* <https://wandb.ai/site>. Reporte los mejores valores encontrados.
 - b) Realice lo anterior para el algoritmo del descenso del gradiente.
 - c) ¿Porqué el algoritmo RMS prop es más efectivo en evitar atascarse en puntos silla que el algoritmo del descenso del gradiente?
 - d) Para los mejores valores encontrados, en cada funcion, por un máximo de 25 iteraciones (para el algoritmo del descenso del gradiente, y RMS prop):
 - 1) Ejecute el algoritmo 10 corridas diferentes, y documente en una tabla la cantidad de iteraciones para converger a el o los puntos minimos (indique si convergió).
 - 2) Muestre el valor promedio de la función minimizada para las 10 corridas, y la cantidad de iteraciones promedio en converger.
 - e) Muestre los puntos visitados del algoritmo para la mejor corrida (convergencia mas rapida) para cada función. Para ello use el gráfico de las curvas de nivel. Además para tales corridas, grafique la curva de aprendizaje.
 - f) Realice una comparativa entre el algoritmo del descenso del gradiente y el algoritmo RMS prop. según los resultados obtenidos.
3. **(30 puntos)** Implemente el algoritmo de **simulated annealing** en pytorch, de la forma mas vectorial posible.
 - a) Realice un proceso de calibración de sus hiper-parámetros y reporte los mejores valores encontrados, usando como evidencia las gráficas de aprendizaje de los mejores hiper-parámetros encontrados.
 - b) Para los mejores valores encontrados, en cada funcion, por un máximo de 25 iteraciones (para el algoritmo del descenso del gradiente, y el descenso del gradiente adaptativo):
 - 1) Ejecute el algoritmo 10 corridas diferentes, y documente en una tabla la cantidad de iteraciones para converger a el o los puntos

minimos (indique si convergió). Los **10 puntos iniciales escogidos deben preservarse para todos los algoritmos a probar** (deben ser los mismos que el **descenso del gradiente** y el **descenso del gradiente adaptativo**).

- 2) Muestre el valor promedio de la función minimizada para las 10 corridas, y la cantidad de iteraciones promedio en converger.
 - c) Muestre los puntos visitados del algoritmo para la mejor corrida (convergencia mas rapida) para cada función. Para ello use el gráfico de las curvas de nivel. Además para tales corridas, grafique la curva de aprendizaje.
 - d) ¿Cómo podría mezclar el algoritmo del descenso del gradiente con el algoritmo de *simulated annealing*? ¿Qué beneficios tendría el hacerlo?
4. **(10 puntos)** Compare los resultados de todos los algoritmos probados, y argumente las ventajas y desventajas de cada uno, usando fuentes externas debidamente citadas.