

# Introducción a TensorFlow

Minería de Datos

Sergio Salomón García

Universidad de Cantabria



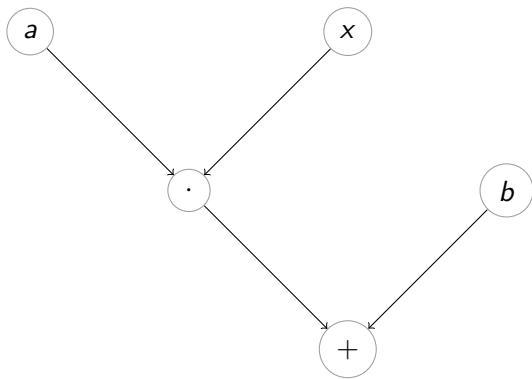
# Contenido

- ▶ Introducción
- ▶ Paradigma de TensorFlow
- ▶ TensorBoard
- ▶ Práctica
  - ▶ Elementos básicos
  - ▶ Ejemplo de regresión lineal
  - ▶ Ejemplo de clustering
  - ▶ Ejemplo de clasificación y softmax
- ▶ Aporte de TensorFlow

# Introducción

- ▶ Librería para computación numérica usando “grafos de flujo de datos” (*data-flow graphs*).
- ▶ Desarrollado originalmente por investigadores e ingenieros de Google Brain Team.
- ▶ Predecesor de DistBelief.
- ▶ Se puede utilizar con los lenguajes Python, C y C++.
- ▶ Facilita el uso de una o varias CPUs/GPUs, y es multiplataforma.
- ▶ Válido para otras tareas más allá del aprendizaje automático.

## Paradigma: grafos de flujo de datos



$$a \cdot x + b$$

# Paradigma: grafos de flujo de datos

- ▶ Las cálculos se construyen como grafos dirigidos.
- ▶ Los nodos representan operaciones matemáticas (*ops*) principalmente.
  - ▶ También entrada/salida o asignación de variables.
- ▶ Los datos deben representarse como *tensores* (arrays multidimensionales).
- ▶ Un nodo toma 0 o más *tensores* de entrada y devuelve 0 o más como salida.
  - ▶ Dicho de otra forma: en las aristas se transmiten siempre tensores.

## Paradigma: ejecución

- ▶ Tras la construcción del modelo, debe lanzarse la ejecución para obtener los resultados.
- ▶ Los nodos deben ejecutarse dentro de una *sesión*.
- ▶ La *sesión* se encarga de distribuir los cálculos del grafo (los nodos) entre los *dispositivos* (hardware) disponibles.
- ▶ La ejecución se realiza de forma asíncrona y en paralelo una vez las aristas de entrada están listas.
- ▶ Se puede obtener (*fetch*) el valor de un nodo mediante `eval` en el mismo o `run` en la sesión.

## Paradigma: elementos

- ▶ **Tensors:** arrays multidimensionales con *rank* (dimensión), *shape* (tamaño) y *dtype* (tipo de datos).
- ▶ **Constants:** tensores con un valor fijo asignado en la creación.
- ▶ **Variables:** contienen tensores y permiten reasignar valores durante la ejecución.
- ▶ **Placeholders:** similar a la constante, pero siempre debe inicializarse (*feed*) al lanzar la ejecución (cada vez).

# TensorBoard

- ▶ TensorFlow incluye una herramienta de visualización interactiva llamada TensorBoard.
- ▶ En esta pueden cargar y mostrarse los grafos contruidos, las ejecuciones y los resultados.



# TensorBoard

[EVENTS](#)[IMAGES](#)[GRAPH](#)[HISTOGRAMS](#)

Fit to screen

Run

/Users/sergio/Desktop/t...

Upload

Choose File

Color

Structure

color: same substructure  
gray: unique substructure

Graph

(\* = expandable)



Namespace\*



OpNode



Unconnected series\*



Connected series\*



Constant



Summary



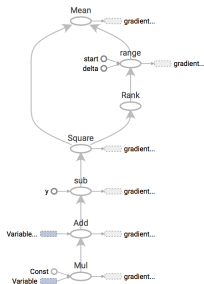
Dataflow edge



Control dependency edge



Reference edge



# Aportes de TensorFlow

- ▶ Los grafos de flujo de datos son altamente paralelizables.
- ▶ Separación de la construcción del grafo (modelo) de la ejecución (o el entrenamiento).
- ▶ Impulsado por Google, abierto a la comunidad y con buena documentación.
- ▶ No se trata de las librerías más rápidas.
- ▶ Aún se encuentra “empezando”.

