Some slides on Jax

STATS 810
Ed Ionides

December 1, 2025

Users

- "JAX is used by almost every large genAI player (Anthropic, Cohere, DeepMind, Midjourney, Character.ai, XAi, Apple, etc.). Its actual market share in foundation models development is something like 80%."
  (https://news.ycombinator.com/item?id=39876444)
- Used for AlphaFold, Google Maps routing, etc.
- Becoming dominant for scientific computing with GPUs

Approach

- Leverages XLA
- Looks like numpy

Thinking like C

- For just-in-time (jit) compilation, Jax needs to know the size of arrays. Any function called with a new array size will be recompiled.
- Loops of variable length are problematic for automatic differentiation (AD). Reverse mode AD needs access to the full computation tree, and so jit wants this to have fixed dimension.

Thinking like Jax

- Central ideas are the composable functions `jax.jit`, `jax.vmap`, `jax.grad`
- `jax.jit`. Just-in-time compilation, building a computation tree.
- `jax.vmap`. Announces a vectorized calculation. This does not specify how the calculation will be split up (**sharded**) across processors.
- `jax.grad`. Uses reverse-mode automatic differentiation. Repeated to give the Hessian.

GPU considerations

- 32 bit vs 64 bit numbers
- To what extent does a GPU follow the single instruction multiple data (SIMD) paradigm?