Northwestern
INFORMATION TECHNOLOGY

Julia

Sophie van Genderen, Assoc Computational Specialist
Julia Giannini, Computational Specialist
Matthew Gorby, Senior Computational Specialist
Northwestern IT Research Computing and Data Services

December 5th 2024

Jupyter Notebooks to Load

1. In Google CoLab, open from Github
2. Go to nuitrcs/Julia_workshop
3. Notebooks to open:
    - mnist_perceptron_training.ipynb
    - Julia_DS_Example.ipynb

4. For answers:
    - mnist_perceptron_training_answers.ipynb
    - Julia_DS_Example_Answers.ipynb

5. Extra machine learning example (on a GPU) for reference:
    - transformer_text_analysis_on_gpu.ipynb

Julia

Topics Covered in this Workshop

- Review Quiz
- Julia usage in 'the real world'
- Data Science examples and packages
- Machine Learning examples and packages
- Working Examples

Review: comparing Python to Julia

Key Indicator	Julia	Python
Maturity	Created in 2012	Created in 1991
Scope	General-purpose, but data-oriented	General-purpose and used for almost everything
Language Type	High Level, (Just in Time) Compiled	High Level, Interpreted
Typing	Dynamically-typed language, but also offers the ability to specify types (Static)	Dynamic, the type for a variable is decided at runtime
Open-source	Yes	Yes
Usage	Data Science and Machine Learning – especially work with data models	Mobile/web Dev, AI, Data Science, web scripting, game development, security ops.
Data Science	Math functions are easy to write and understand – no external libraries are needed for math functions	Requires NumPy or other libraries for advanced math
Performance	Fast development and production, high speed runtime, can handle millions of data threads	Fast for development, slow for production

What Companies Use Julia

- BlackRock
- NASA
- Microsoft
- Pfizer
- IBM
- Amazon
- JP Morgan AI Research

Companies like **BlackRock, NASA, Microsoft, Pfizer, IBM, Amazon, JP Morgan AI Research, and ASML** are known to use Julia for coding, primarily due to its high performance for numerical computing and data analysis tasks across various sectors like finance, aerospace, and pharmaceuticals.

Key points about Julia usage in companies:

Financial institutions:

BlackRock heavily utilizes Julia for quantitative analysis and modeling due to its speed and ease of use.

Research and development:

NASA employs Julia for large-scale data analysis and simulations in areas like planetary science.

Pharmaceutical industry:

Companies like Pfizer use Julia to accelerate drug discovery simulations.

Technology giants:

Microsoft, IBM, and Amazon have adopted Julia for specific projects within their organizations.

Data Science

Packages

- https://juliapackages.com/c/data-science

Machine Learning

Why Julia

- https://www.datacamp.com/blog/introduction-machine-learning-julia

Machine Learning packages

- **Notebook**: Pluto, IJulia, Jupyter
- **Package/environment management**: Pkg
- **Importing and handling data**: CSV, DataFrames
- **Plotting and output**: Plots, StatsPlots, LaTeXStrings, Measures, Makie
- **Statistics and Math**: Random, Statistics, LinearAlgebra, StatsBase, Distributions, HypothesisTests, KernelDensity, Lasso, Combinatorics, SpecialFunctions, Roots
- **Individual machine learning packages**:
  - Generalized linear models (e.g. linear regression, logistic regression): GLM
  - Deep Learning: Flux, Knet
  - Support vector machines: LIBSVM
  - Decision tree, random forest, AdaBoost: DecisionTree
  - K-nearest neighbors: NearestNeighbors
  - K-means clustering: Clustering
  - Principal component analysis: MultivariateStats

Machine learning examples

- Transformer language processing:
  - https://juliapackages.com/p/transformers
  - https://developers.google.com/machine-learning/clustering/dnn-clustering/supervised-similarity

- MNIST dataset:
  - https://github.com/FluxML/model-zoo/tree/master/vision/mlp_mnist
  - https://github.com/FluxML/model-zoo/tree/master?tab=readme-ov-file#examples-in-the-model-zoo
  - https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi

More machine learning and data science examples in Julia

- Time series data / Fourier analysis
  - https://github.com/Marco-Congedo/FourierAnalysis.jl
- Iris flower example / Clustering
  - https://github.com/jbrea/MLCourse/blob/main/notebooks/clustering.jl

Quest – Northwestern's Super Computer

You'll need a classroom or research allocation to access Quest's compute resources – more information here: https://www.it.northwestern.edu/departments/it-services-support/research/computing/quest/general-access-allocation-types.html

Quest User Guide: https://services.northwestern.edu/TDClient/30/Portal/KB/ArticleDet?ID=505

Quest OnDemand – helpful user interface for interactive applications: https://services.northwestern.edu/TDClient/30/Portal/KB/ArticleDet?ID=2234

Research computing how-to videos: https://services.northwestern.edu/TDClient/30/Portal/KB/ArticleDet?ID=2004

Thank You!

Questions about Quest? Email us at: quest-help@northwestern.edu