

DynaML: A Scala Library/REPL for Machine Learning Research

Mandar H. Chandorkar^{1, 2}

1 Centrum Wiskunde en Informatica, Multiscale Dynamics 2 INRIA Paris-Saclay/Laboratoire de Recherche en Informatique, TAU

DOI: [10.21105/joss.00802](https://doi.org/10.21105/joss.00802)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 30 June 2018

Published: 01 July 2018

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

DynaML is a Scala platform which aims to provide the user with an toolbox for research in data science and machine learning. It can be used as

- A scala shell, local or remotely hosted.

```
$ dynaml
```

```
Welcome to DynaML v1.5.3-beta.3
```

```
Interactive Scala shell for Machine Learning Research
```

```
Currently running on:
```

```
(Scala 2.11.8 Java 1.8.0_101)
```

```
DynaML>
```

- A standalone script engine.

```
$ dynaml ./scripts/cifar.sc
```

- As a binary dependency for JVM based machine learning applications.

```
libraryDependencies += "com.github.transcendent-ai-labs" % "DynaML" % "master-SNAPSHOT"
```

Motivation & Design

DynaML aims to provide an *end to end* solution for research and development in machine learning, statistical inference and data science. Towards these goals, it provides the user with modules for.

- Data pre-processing using functional transformations. These transformations or *pipes*, can be joined to form complex processing pipelines.
- Training [predictive models](#), with a special focus on *stochastic processes*, *kernel* methods & neural networks. The [model API](#) can be extended to implement customized and complicated algorithms.
- [Model tuning](#) & hyper-parameter optimization.
- [Model evaluation](#)
- Visualization: two and [three](#) dimensional charts.

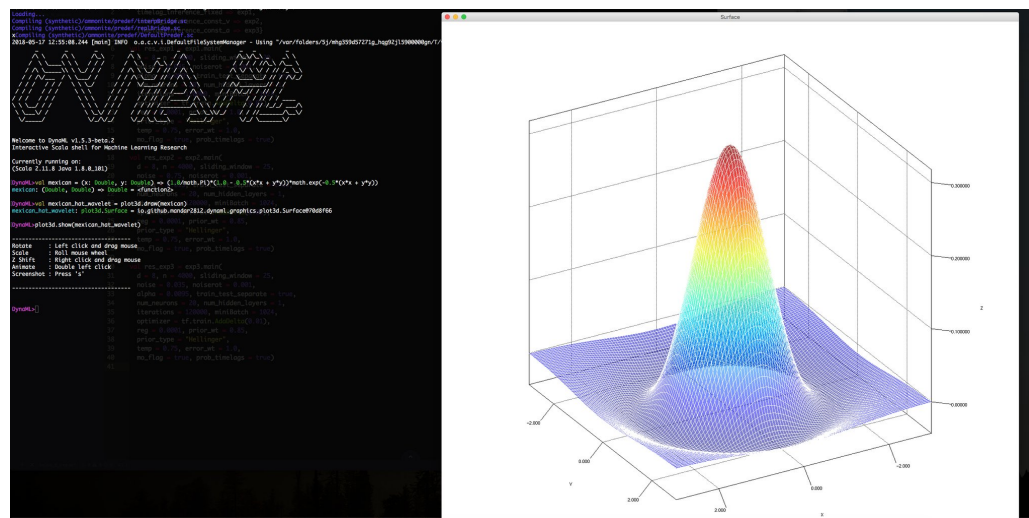


Figure 1: Example figure.

Scala Ecosystem

Scala (Scala Center 2002) is a high level *object oriented & functional* programming language which runs on the *Java Virtual Machine* (JVM). Its expressiveness, multi-threading model and ability to execute on the JVM enable the prototyping and development of potentially large scale and data intensive applications.

The scala eco-system has a number of useful packages which DynaML leverages such as, Tensorflow (???) support through *Tensorflow for Scala* (Platanios 2017), the [breeze](#) linear algebra library and the [Ammonite](#) project.

Applications

DynaML has been applied in research into *Gaussian Process* based geomagnetic time series prediction (Chandorkar, Camporeale, and Wing 2017) & (Chandorkar, Camporeale, and Wing 2017) and in on-going research in MCMC based Bayesian inverse PDE problems specifically *Fokker Planck* based plasma radial diffusion systems (Camporeale and Chandorkar 2017).

It can be accessed via the [online repository](#), or imported as a managed dependency into JVM projects via [jitpack](#).

The [user guide](#) contains information regarding installation, usage, API documentation (Scaladoc) as well as usage examples.

Acknowledgements

DynaML was conceived during the [Master of Science, Artificial Intelligence](#) program at the KU Leuven and further developed during the PhD research carried out in the project [Machine Learning for Space Weather](#) which is a part of the [CWI-INRIA International Lab](#).

References

- Camporeale, E., and M. Chandorkar. 2017. “Bayesian inference of radiation belt loss timescales.” *AGU Fall Meeting Abstracts*, December.
- Chandorkar, M., E. Camporeale, and S. Wing. 2017. “Probabilistic Forecasting of the Disturbance Storm Time Index: An Autoregressive Gaussian Process Approach.” *Space Weather* 15 (8):1004–19. <https://doi.org/10.1002/2017SW001627>.
- Platanios, Emmanuel Antonios. 2017. “Tensorflow Scala: Scala Api for Tensorflow.” http://platanios.org/tensorflow_scala/.
- Scala Center, École Polytechnique Fédérale Lausanne (EPFL). 2002. “Scala: The Scala Programming Language.” <https://www.scala-lang.org>.