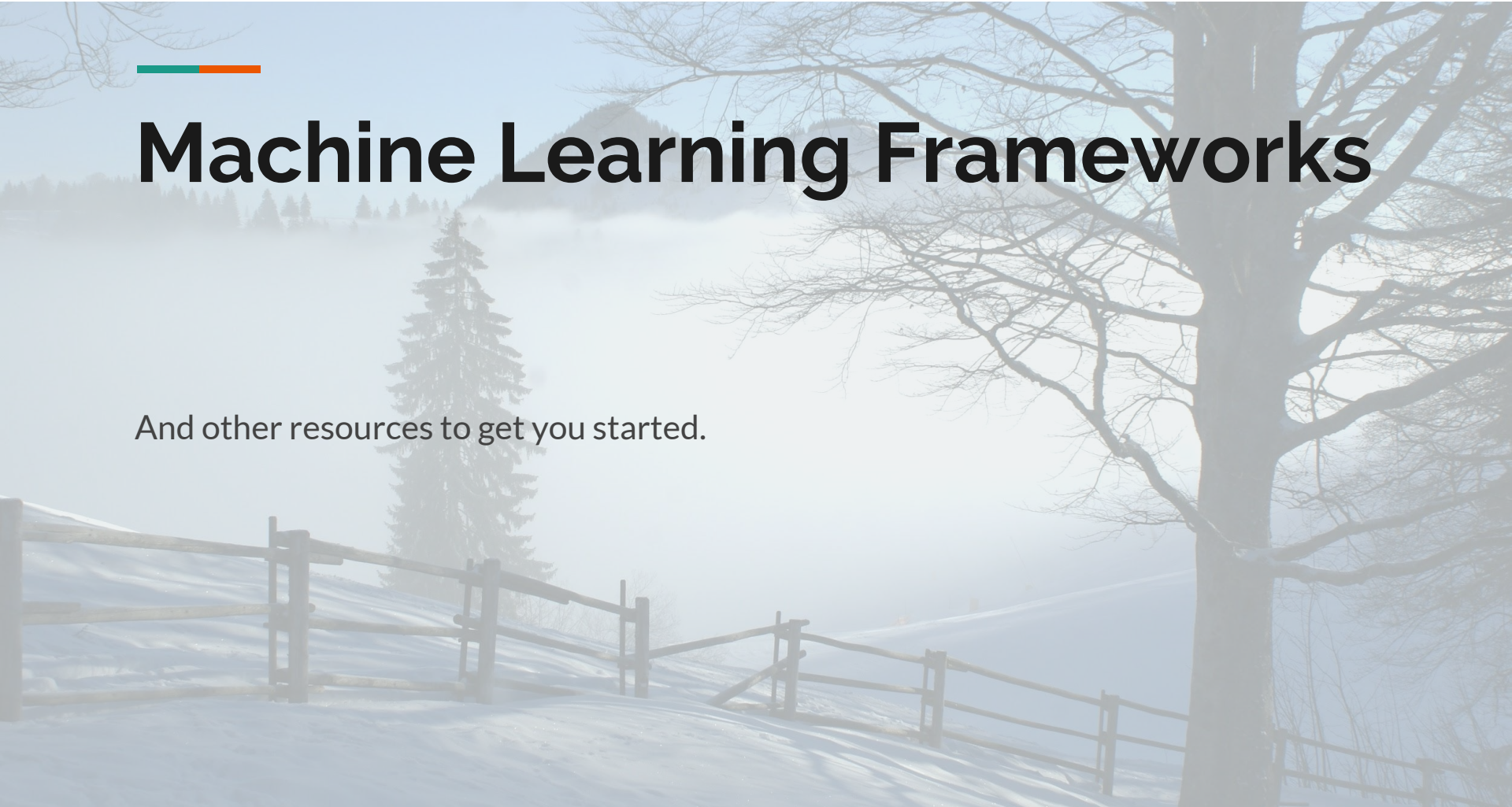
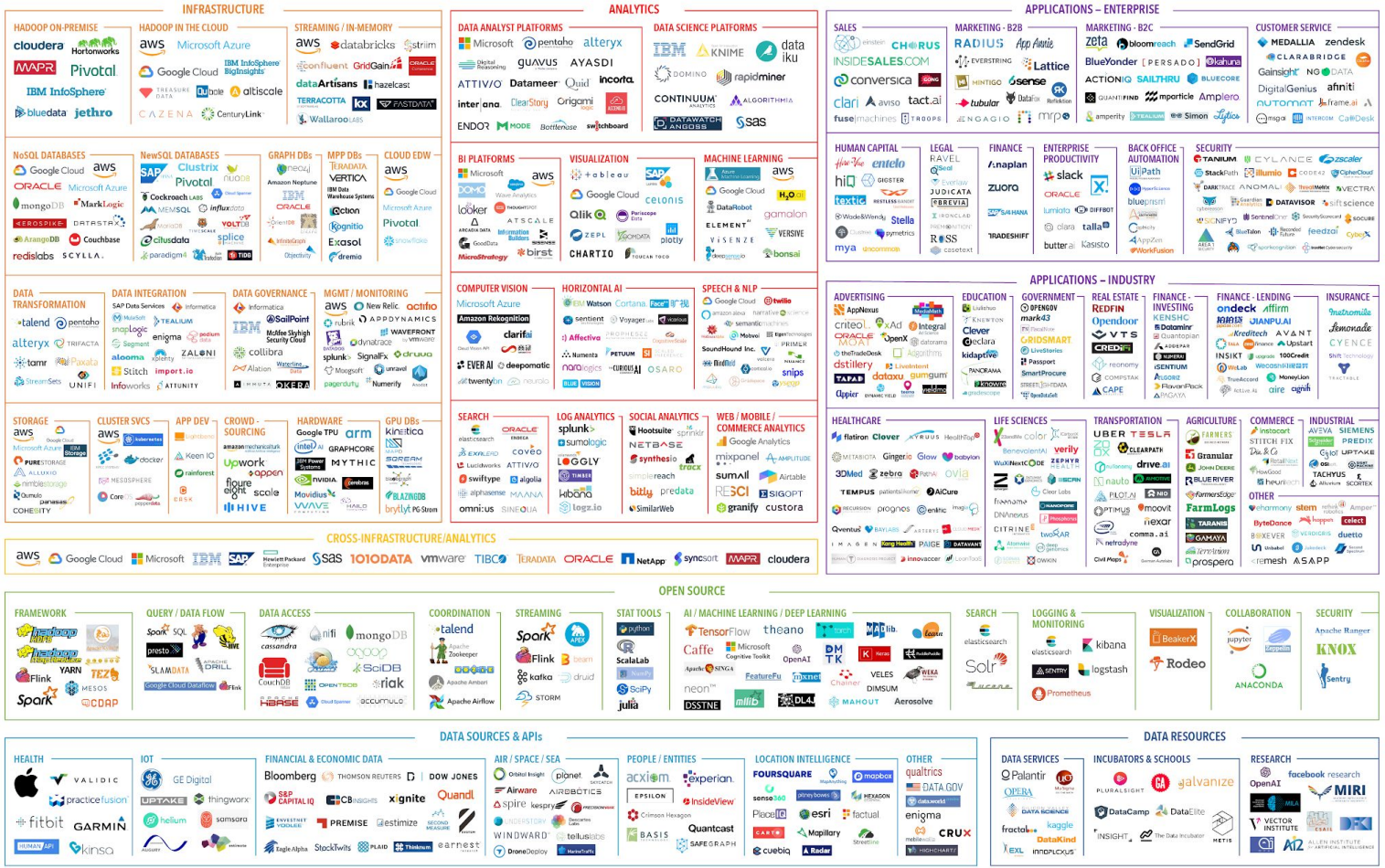


Machine Learning Frameworks

And other resources to get you started.

BIG DATA & AI LANDSCAPE 2018



FIRSTMARK
EARLY STAGE VENTURE CAPITAL

STAT TOOLS

ScalaLab

AI / MACHINE LEARNING / DEEP LEARNING

VELES

DIMSUM

We'll focus on Python here.

Best for Classical ML Algorithms

- Python library built on Numpy, SciPy, Matplotlib
- Supports most of the classical supervised and unsupervised algorithms: *linear and logistic regressions, SVM, Naive Bayes, etc.*
- Easy to read *[arguably]*
- Limited support for Neural Networks; i.e. can't be used for Deep Learning
- Has datasets ready to play around with; e.g. MNIST, Iris flower, housing prices
- Huge community

XGBoost, LightGBM and CatBoost: packages for gradient boosting over decision trees algorithms

The most popular Machine Learning and Deep Learning Library

Popular for Neural Networks and Deep Neural Networks

Offers a tool to visualize your ML models: TensorBoard

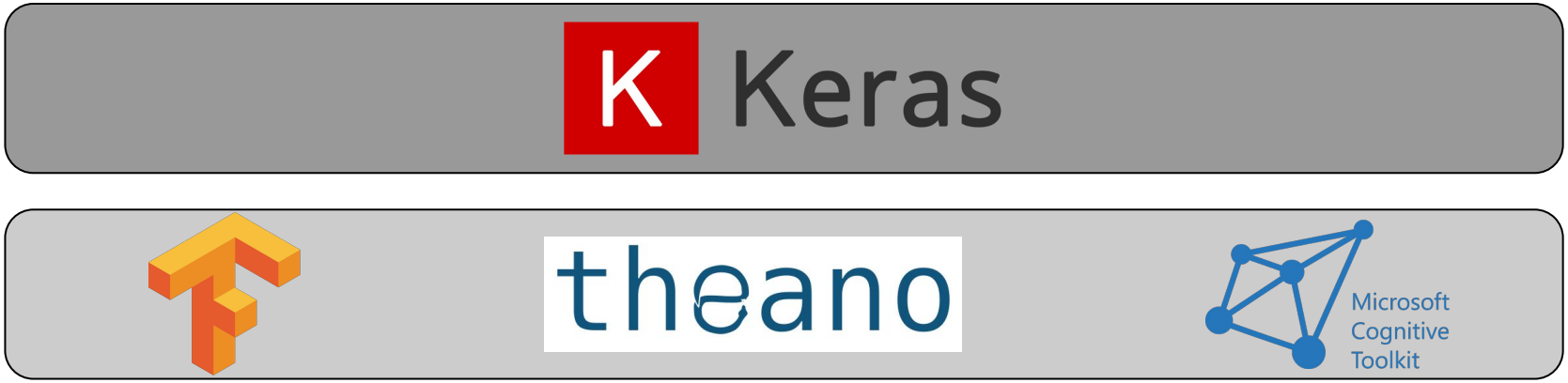
A bit of a learning curve...

No pretrained models

Must define the entire computational graph of your model before running

Can be deployed on multiple CPUs and GPUs

Keras


Keras
theano
Microsoft
Cognitive
Toolkit
CUDA/cuDNN
BLAS, Eigen
GPU
CPU

Keras

- Adds a lot of convenience to for example tensorflow
- Build up your neural network in building blocks
- You have been seen it a bunch during this week

TensorFlow’s less popular cousin...

theano

At its core it is an auto-differentiation framework.

Simpler to use and faster than TensorFlow when it comes to usability and speed

Raw theano is somewhat low-level

Error messages are not too helpful

Can only be deployed on a single GPU

Not fully maintained anymore (basically only maintained to the level that PyMC needs it). Use at your own risk.

Another Deep Learning Library

‘the new kid on the block’

More Pythonic and flexible than TensorFlow

Less ‘mature’ than TensorFlow, but the support and community is growing

Also an auto-differentiation framework at its core

Supports CPU and GPU - accelerated computations

Very popular in research today

Ok, so...which do I use?

Comparisons:

Software	Open Source	Platform	Interface	OpenMP Support	Parallel	Used for
Keras	Yes	Linux, macOS, Windows	Python, R	If using Theano backend		DL
Matlab + NN Toolbox	No	Linux, macOS, Windows	MATLAB	No		DL
Microsoft CNTK	Yes	Linux, Windows	Python, C++, CLI	Yes		DL
PyTorch	Yes	Linux, Windows, macOS	Python	Yes		DL
ONNX	Yes					DL
Scikit Learn	Yes	Linux, Windows, macOS	Python [Library]			ML
Tensorflow	Yes	Linux, Windows, macOS, Android	Python, C++, Java, Go, R,Julia, Swift,Javascript	No		DL

Want to learn more?

Tutorials and Blogs

Coursera & Other Online Courses

Stack Overflow and Google in general ;)

Kaggle

Textbooks on statistics and machine learning.

Suggested Resources:

Scikit-learn tutorials

An Introduction Book to ML with Jupyter Notebooks: 'Hands-On Machine Learning with Scikit Learn & TensorFlow'

Keras Tutorial: The Ultimate Beginner's Guide to Deep Learning in Python

Stanford's CS231N: http://cs231n.stanford.edu/, including pdf slide sets for more on Convolutional Neural Network

Some Book Suggestions: 'The Elements of Statistical Learning', Deep Learning Textbook....

Python Data Science Handbook

- Tutorial style book to learn how to properly use Python and its data science tools
- Available for free online in its entirety in the form of Jupyter notebooks

https://github.com/jakevdp/PythonDataScienceHandbook