Astroinformatics school - "Rise of the machines"

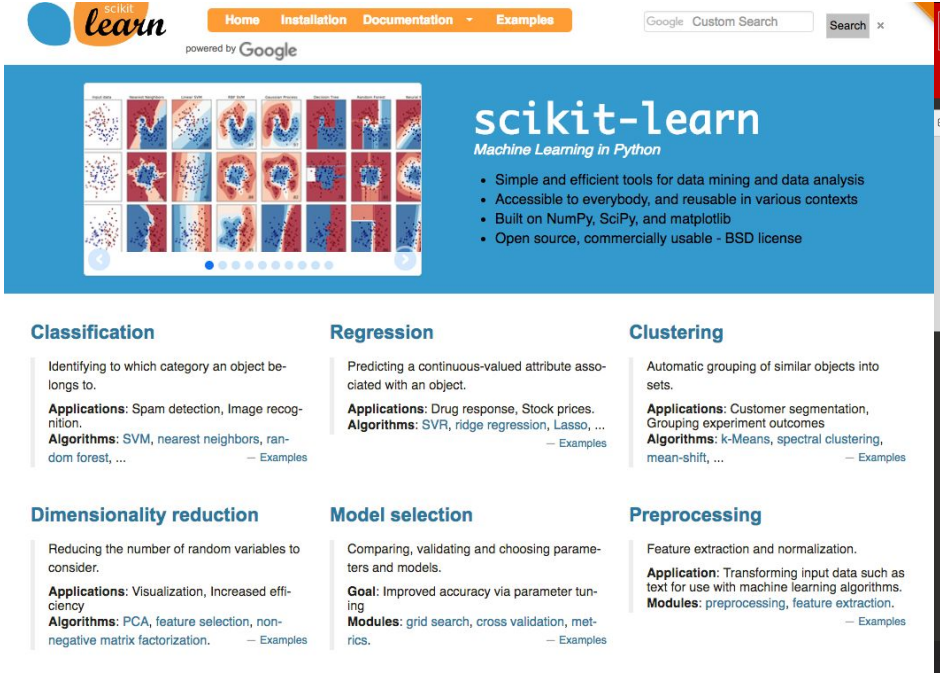
Curtin University
THE CURTIN INSTITUTE FOR COMPUTATION

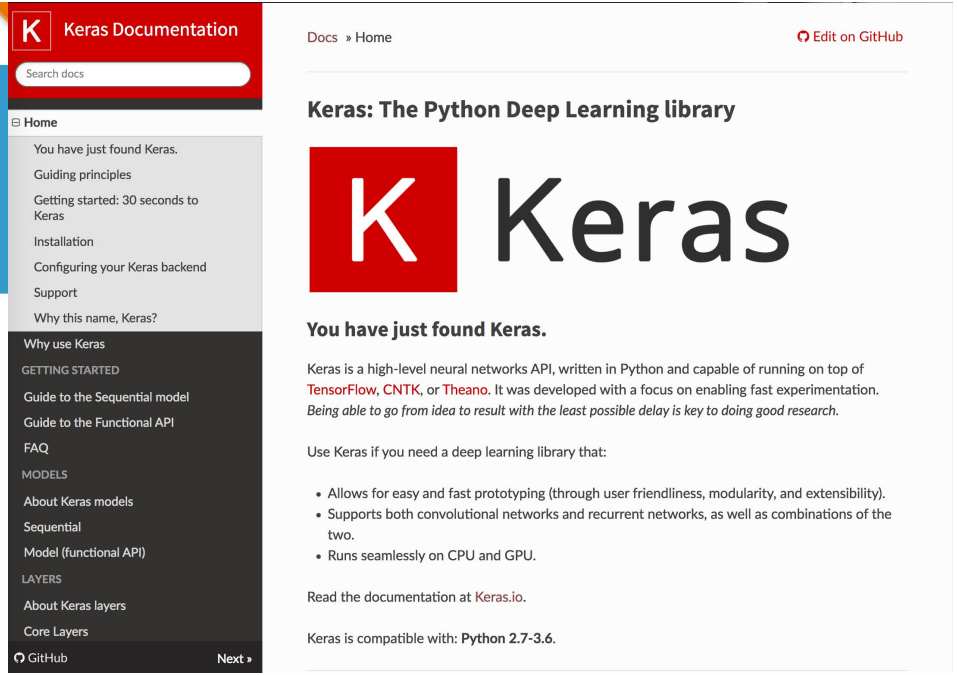
4 to 6 February 2019

Presented by Rebecca Lange and Dan Marrable

Reference sheets

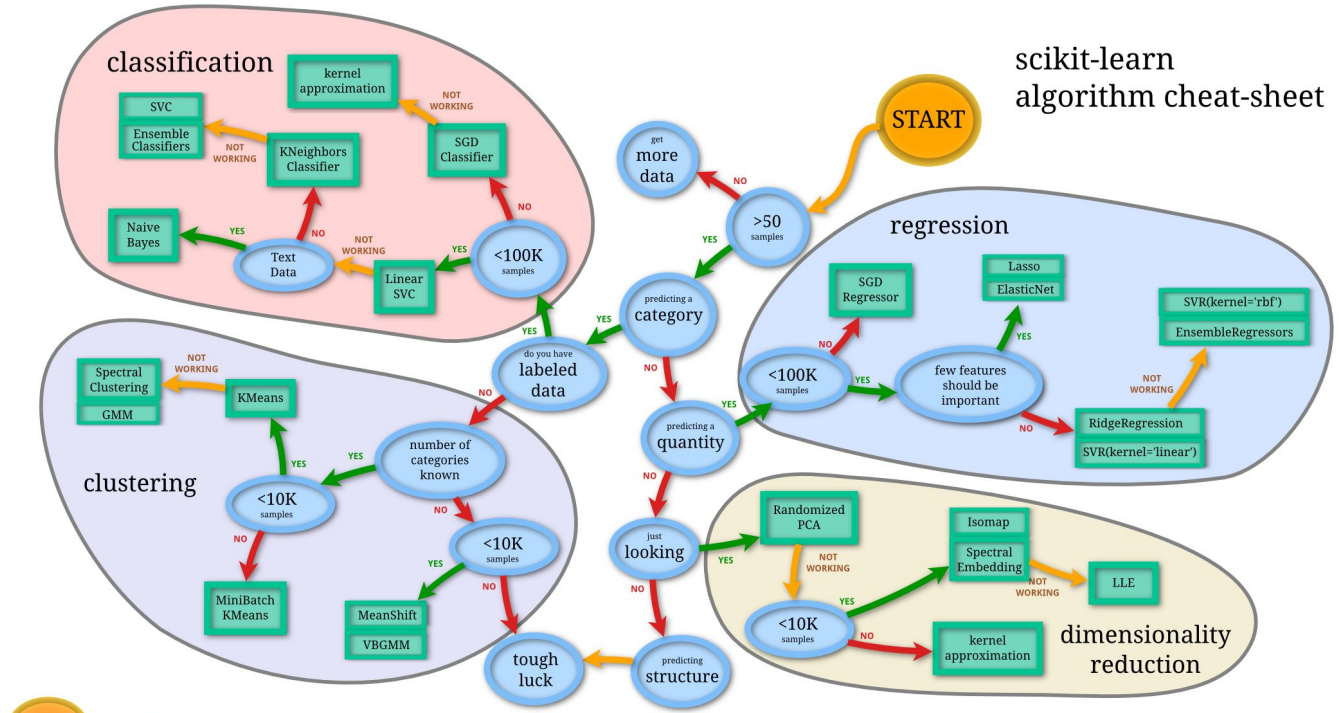
ADACS
ASTRONOMY DATA AND COMPUTING SERVICES





Reference sheets

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES



Back

Source: http://scikit-learn.org/stable/tutorial/machine_learning_map/

Reference sheets

ADACS
ASTRONOMY DATA AND COMPUTING SERVICES

Comparison of supervised machine learning techniques

Factors	Decision Trees	Neural Networks	Naïve Bayes	kNN	SVM	Rule Learners
General accuracy	**	***	*	**	****	**
Learning speed	***	*	****	****	*	**
Classification speed	****	****	****	*	****	****
Tolerance to missing values	***	*	****	*	**	**
Tolerance to irrelevant features	***	*	**	**	****	**
Tolerance to redundant features	**	**	*	**	***	**
Tolerance to highly related features	**	***	*	*	***	**
Dealing with discrete, binary and continuous features	****	***	***	***	**	***
Tolerance to noise	**	**	***	*	**	*
Dealing with model overfitting	**	*	***	***	**	**
Attempts for incremental learning	**	***	****	****	**	*
Explanation of classification	****	*	****	**	*	****
Model parameter handling	***	*	****	***	*	***

* kNN = k-Nearest Neighbours
* SVM = Support Vector Machines

Source: Kotsiantis et al., Supervised machine learning: A review of classification techniques, *Conference on Emerging Artificial Intelligence Applications* 2007.

Reference sheets

ASTRONOMY DATA AND COMPUTING SERVICES

- Supervised learning
  - http://scikit-learn.org/stable/supervised_learning.html#supervised-learning
- Performance metrics
  - http://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics
- Feature selection
  - http://scikit-learn.org/stable/modules/feature_selection.html
- Model selection (cross validation, learning curves, hyperparameter tuning)
  - http://scikit-learn.org/stable/modules/classes.html#module-sklearn.model_selection
- Clustering
  - http://scikit-learn.org/stable/modules/clustering.html#clustering
- Dimensionality reduction
  - http://scikit-learn.org/stable/modules/classes.html#module-sklearn.decomposition
- Deep Learning with Python
  - https://github.com/fchollet/deep-learning-with-python-notebooks

Reference sheets

ASTRONOMY DATA AND COMPUTING SERVICES

- Anomaly detection
  - http://scikit-learn.org/stable/modules/outlier_detection.html
- Time series
  - http://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
  - https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/
  - http://www.statsmodels.org/dev/generated/statsmodels.tsa.arima_model.ARIMA.html
- ML blogs:
  - https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/
  - http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Reference sheets

ASTRONOMY DATA AND COMPUTING SERVICES

- Online courses
  - Machine learning
    - https://www.coursera.org/learn/machine-learning
    - https://www.udacity.com/course/intro-to-machine-learning--ud120
  - Deep learning
    - https://www.coursera.org/specializations/deep-learning
    - https://www.udacity.com/course/deep-learning--ud730
    - http://course.fast.ai/
- “Playgrounds”
  - Tensorflow playground https://playground.tensorflow.org/
  - Visualising the mnist data going through a network http://scs.ryerson.ca/~aharley/vis/