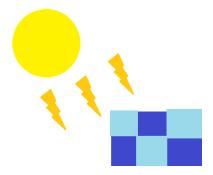# Energy Forecasters
Technology Review

Daniel Chai, MS, Materials Science
Ruidong Ma, MS, Mechanical Engineering
Zang Le, PhD, Chemical Engineering
Xinyue Wang, MS, Materials Science

Background

Topic: Real-time solar power generation estimated from current weather data. Historical weather and energy data used in training.

Machine Learning techniques:

Feature Selection - Random Forest, Recursive Feature Elimination

Predictive Modeling - Neural Network, Multiple Linear Regression

Machine Learning Packages: there's so many …
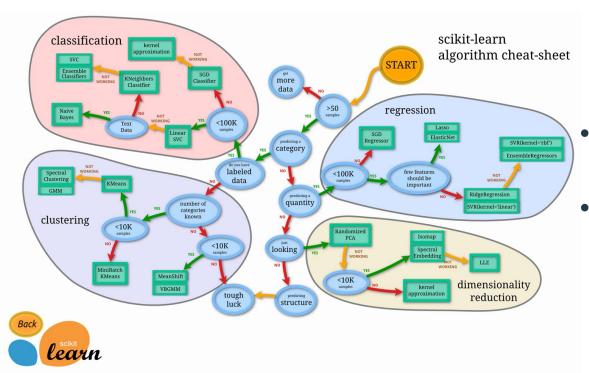
Scikit Learn

What is Scikit learn :

    Open source; Simple and efficient tool for data mining and data analysis

    Classification, regression, clustering, dimensionality reduction, model selection, and preprocessing.

Summary of how it works:

1. Pre-processing: normalization and feature extraction;

2. Classification: random forest (`RandomForestClassifier`) --- Feature importance, Model selection.

3. Prediction : Neural Network, Multiple Linear Regression.

train	validation	test

What's good?

scikit-learn algorithm cheat-sheet

- Almost all ML algorithms can be imported from scikit learn
- Python's Scikit-Learn library can be used to implement the random forest algorithm to solve regression, as well as classification, problems.

Random Forest in scikit learn

1. **Preparing Data For Training:** The first task is to divide data into 'attributes' and 'label' sets. The resultant data is then divided into training and test sets.

        from sklearn.model selection import train_test_split

1. **Training the Algorithm:** The RandomForestRegressor class of the sklearn.ensemble library is used to solve regression problems via random forest.

        from sklearn.ensemble import RandomForestRegressor

1. **Evaluating the Algorithm:** Algorithms are evaluated with mean absolute error, mean squared error, and root mean squared error.

        from sklearn import metrics

6

Drawbacks

Scikit-learn does not offer flexibility for doing supervised neural networks.

- Basics: Feed forward, then back propagation.
- Possibly lower speed. CPU based.

TensorFlow: more options of building networks, allow deep learning, GPU parallel processing.
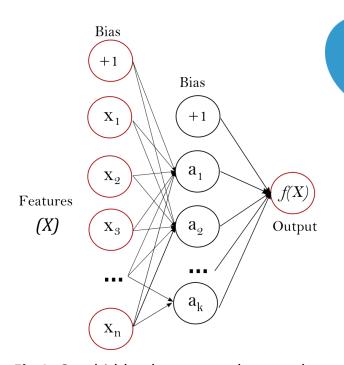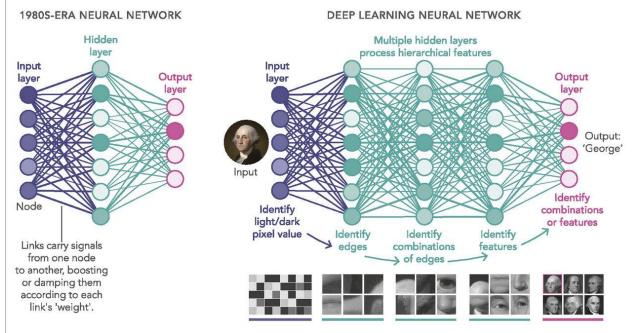
Drawbacks

Fig 1: One hidden layer neural network
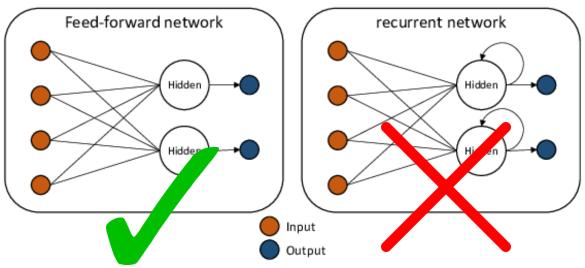
Drawbacks

Drawbacks

Fig 3: Comparison between Feed-forward and recurrent network

Thank you !