Group of Organic Semiconductor

UNIVERSITY *of* WASHINGTON

Overall Project

Goal: Predicting the bandgap and band edge position of organic semiconductor materials

Method: Supervised Machine Learning

Data: HCEPDB_moldata

Task:

1. extract molecular descriptors from SMILES string
2. develop machine learning model
3. Test machine learning model
4. Documentation

Group Members

Yuhuan Meng: Building model

Liang Xu: Determine inputs

Zhi Peng: Testing model

Hongbo Qiao: Documentation

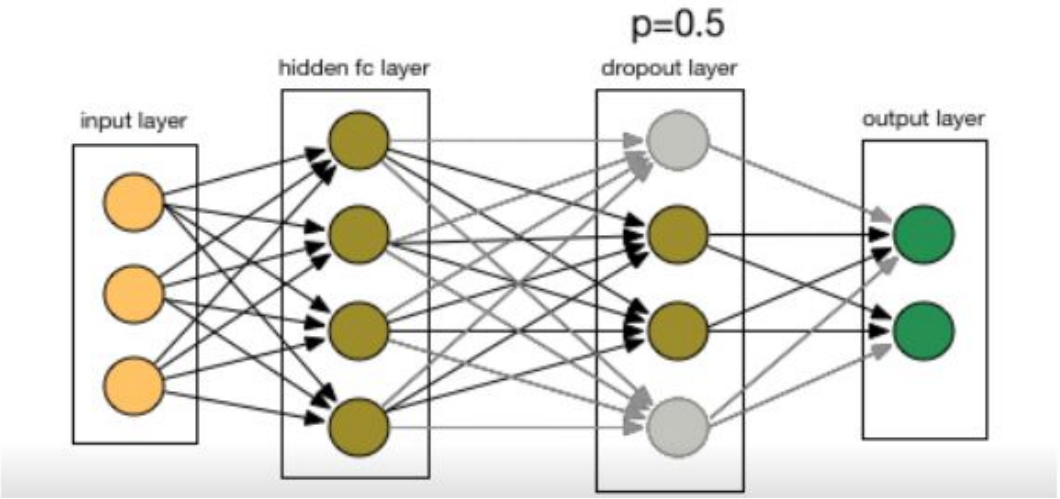
Pre Preparation: Determine inputs

Package:

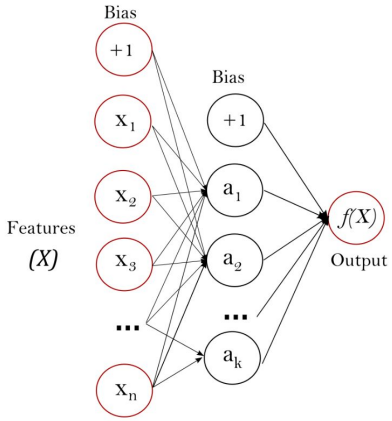
(1) Descriptor Calculator:

	RDkit	PaDEL
Reason for choose	precise calculation results	sometimes error calculation results
Example	C17H14O7 nH count→14	C17H14O7 nH count→12

(2) Regression: Regularization



ML: Neural Network models



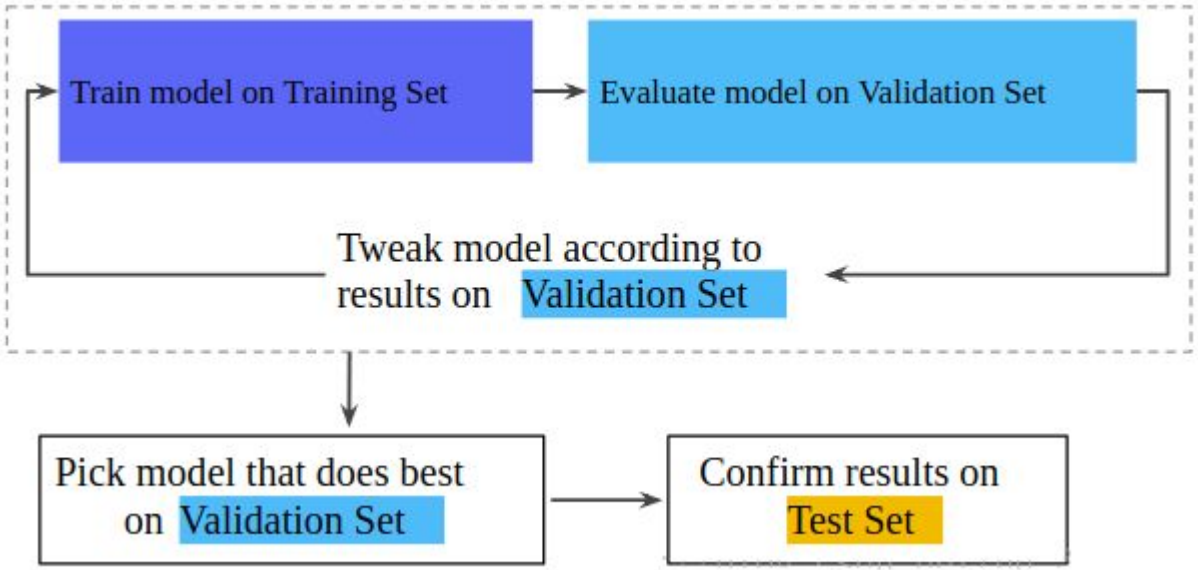
Method: neural network models vs. regression models

Package: Keras vs. scikit-learn

Keras is an library for developing and evaluating deep learning models.
It can build Neural Network at a high-level of the interface.

Main component: neural network sequential model(dense connected layer)
shortcoming: It is hard to build a moderately complex network.

ML: Data Testing



Package: Scikit-learn, Milk

Documentation

To Choose The Editor:

For Markdown(.md):

Jupyter Notebook, Dillinger, Atom...

For Python(.py):

Atom, Idel, Sublime, Jupyter...

Documentation

	Dillinger	Atom
Advantage	1. Completely Online, no install needed. **2. Edit and read at the same time.** 3. Auto save and personal profile.	1. Open Sources, 6000+ packages 2. Free! 3. Customize everything.
Disadvantage	1. Cannot continue to edit the file on another computer.	1. Slower than Sublime in some cases. 2. Need to install packages before use.