MANCHESTER 1824
The University of Manchester

Development of learning workflows for smart synbio design

Ran Duan

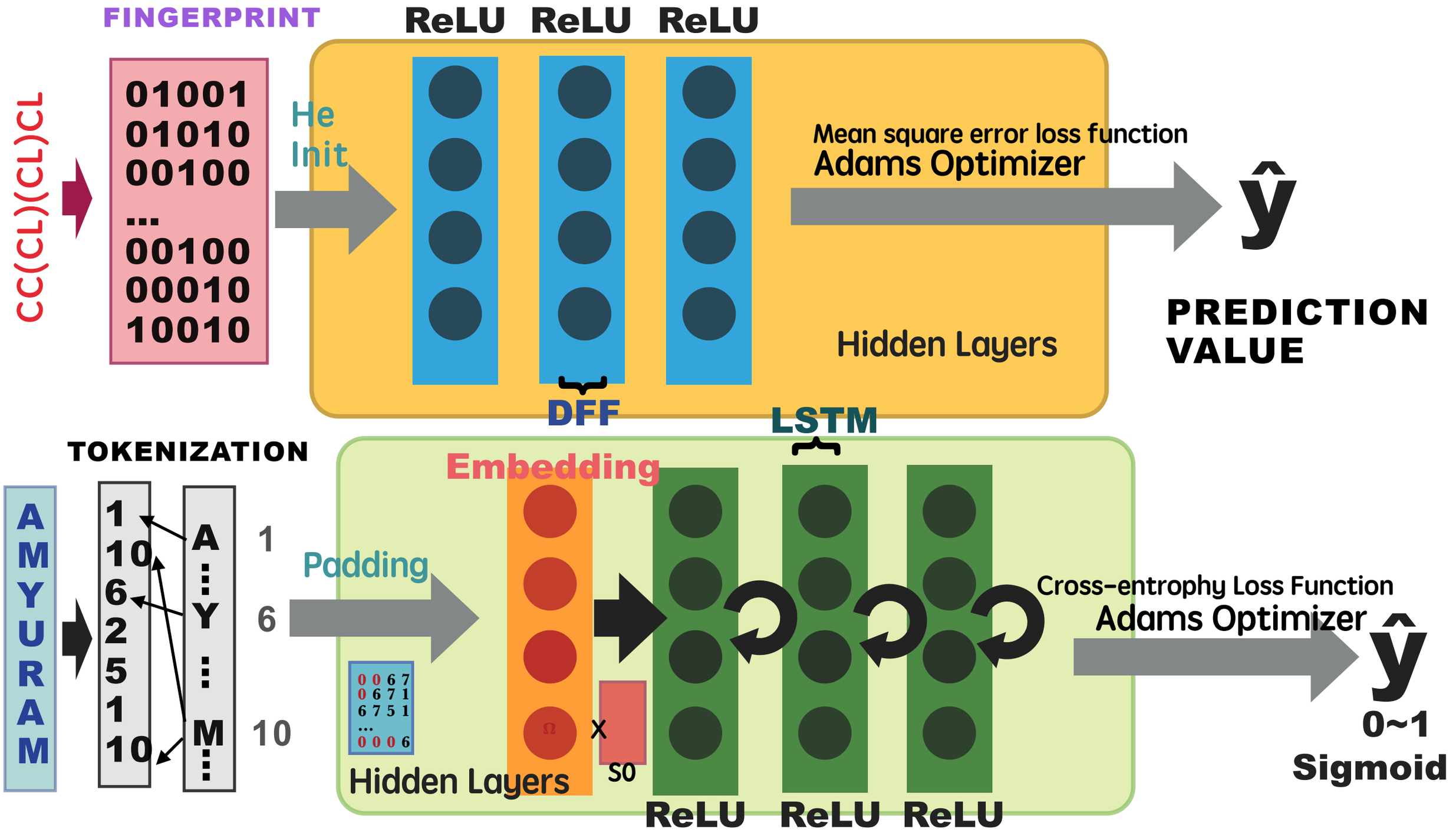
AIMS

Developing machine learning tools linked to experimental protocols at the Learn stage of the Synbio Design/Build/Test/Learn pipeline.

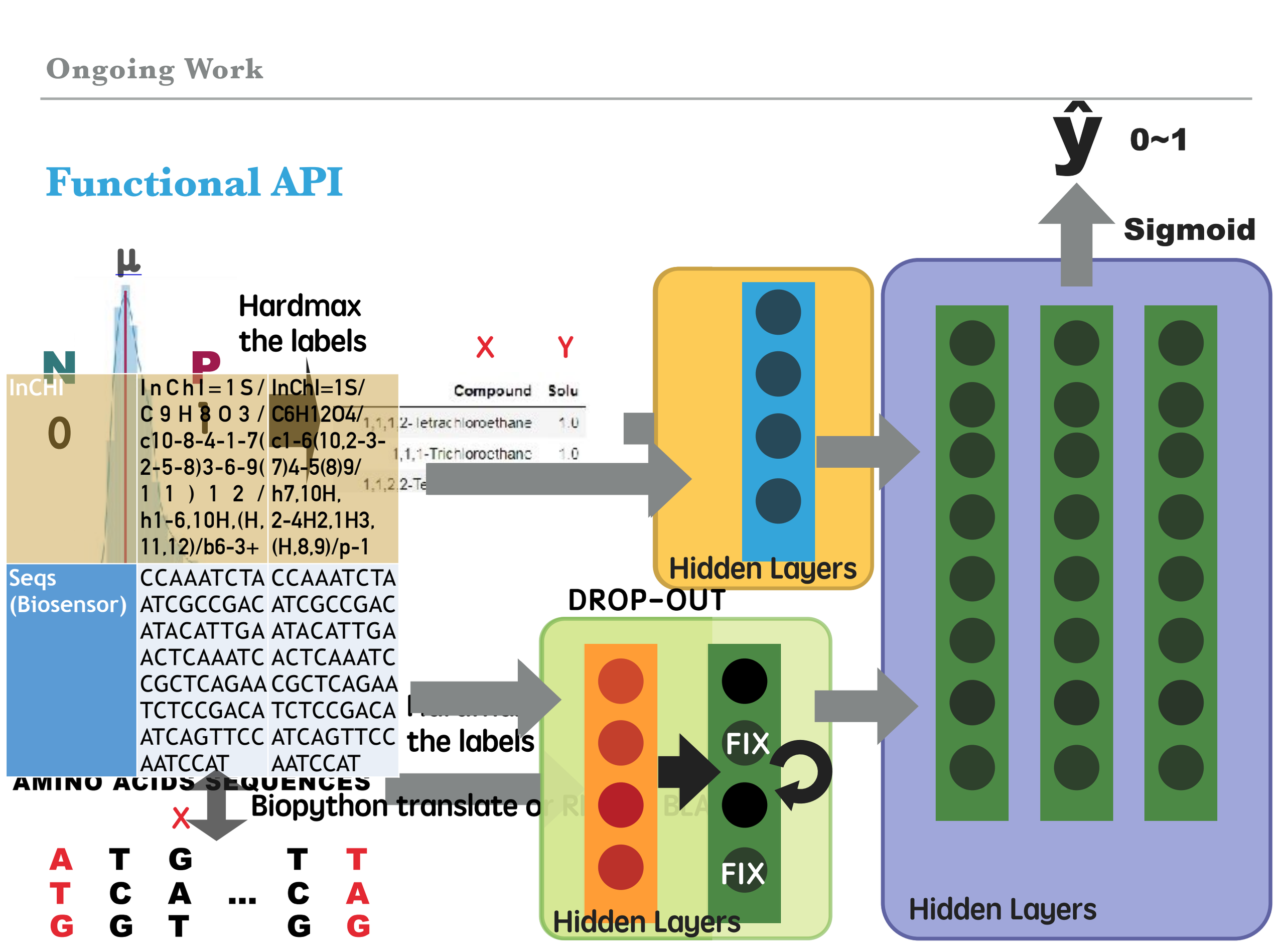
TRANING SET:

- **TF-based biosensors**
- **Sequence reaction scoring**
- **Chemical solubility and sequence thermostability**
- **Synbio parts: RBS, promoters and etc.**

Sequential Neural Networks



Functional API

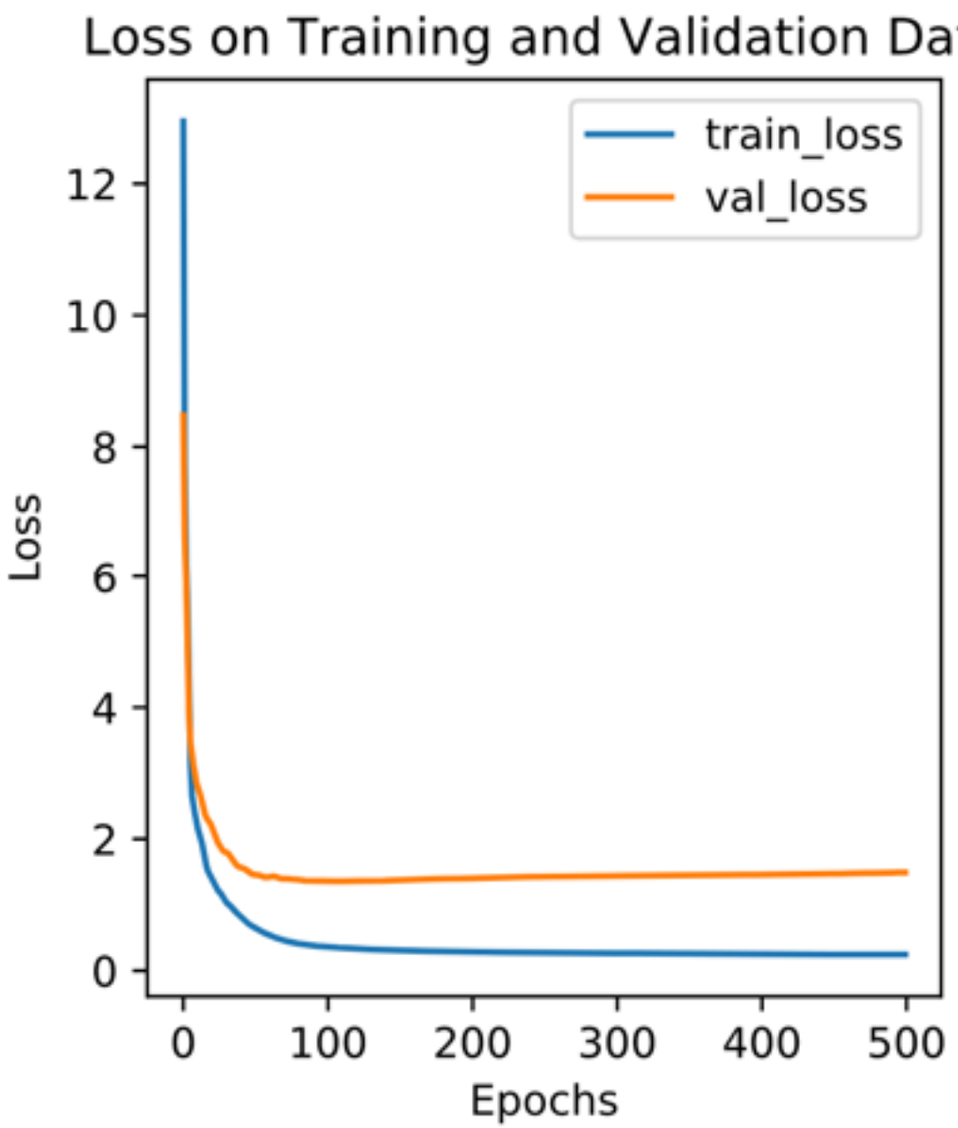


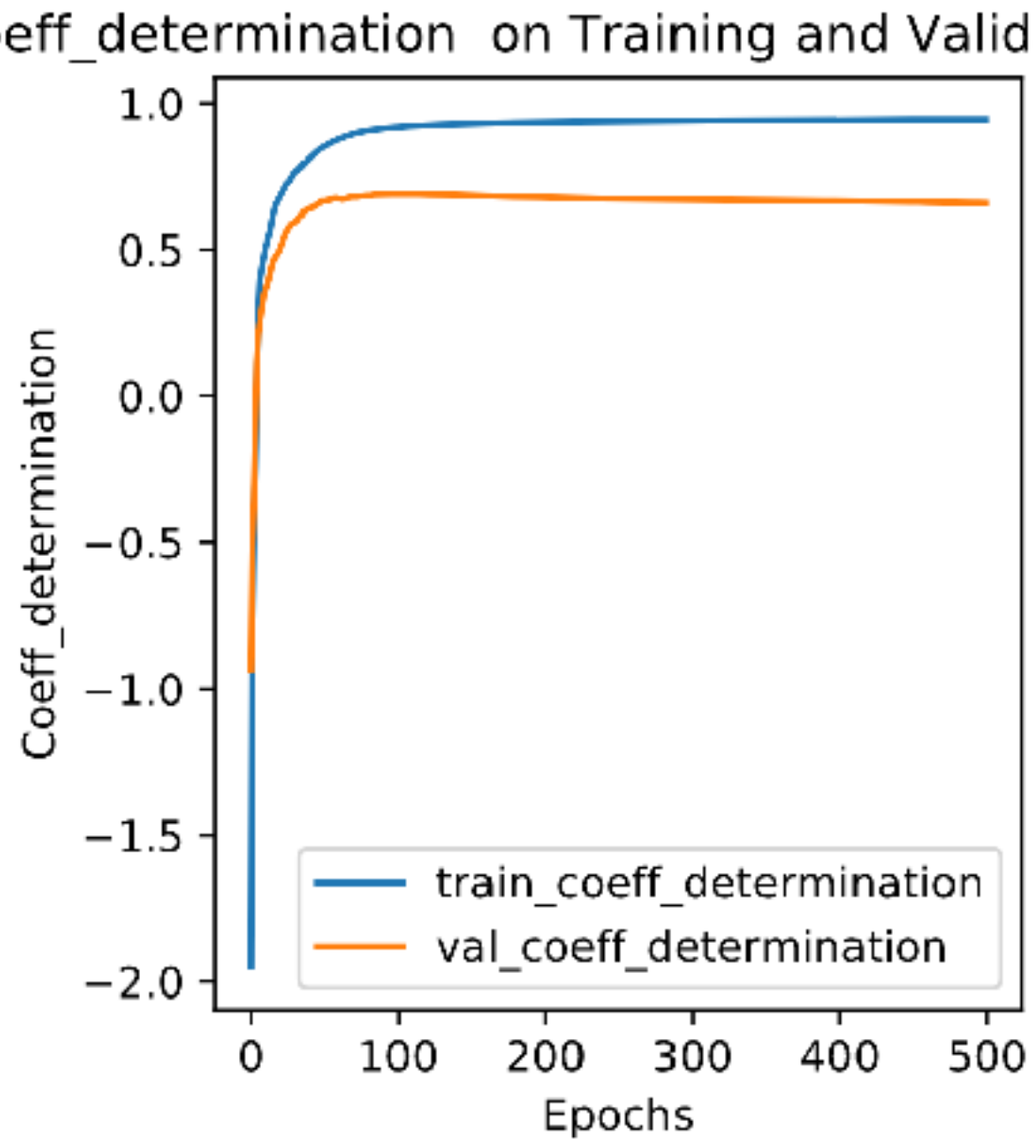
Tuning Initialization Parameters

Solubility prediction

Batch-size: 1144; Epochs: 500	Random Initialization with no bias	Random Initialization	He Initialization
Time (average in 3 folds)	59 us/step	69 us/ step	71 us/ step
R-square (average in 3 folds)	0.9456	0.9307	0.9344
Q-square (average in 3 folds)	0.6164	0.6022	0.6435

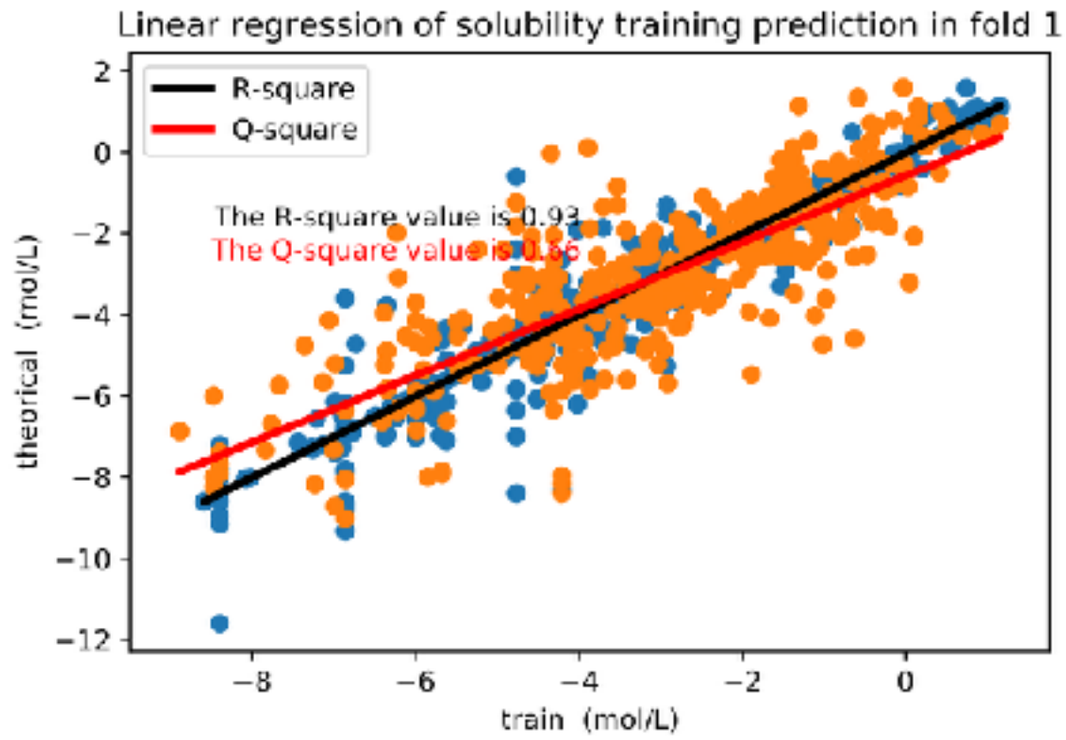
Training Performance

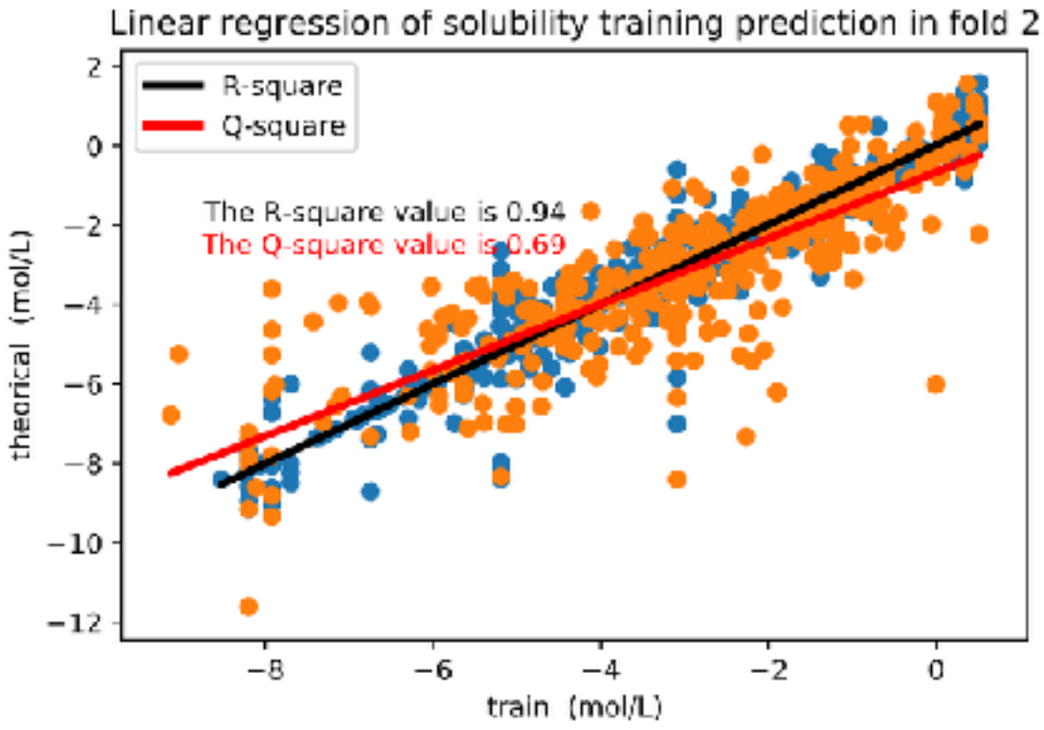


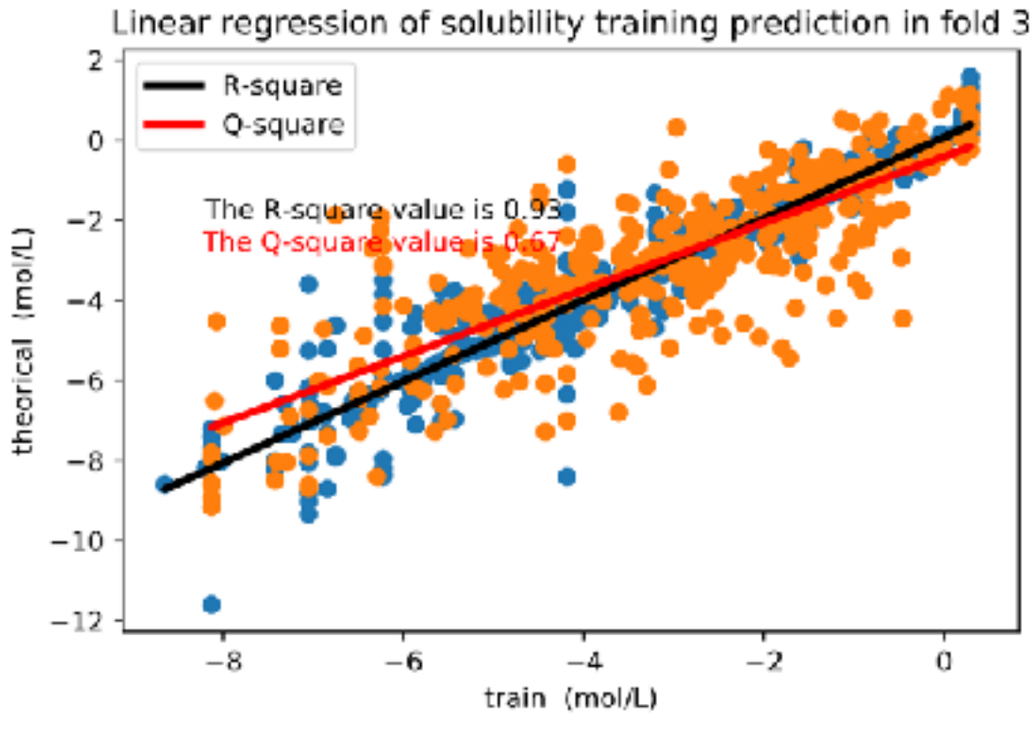


Linear Regression

- Observation and prediction values in testing set
- Observation and prediction values in training set







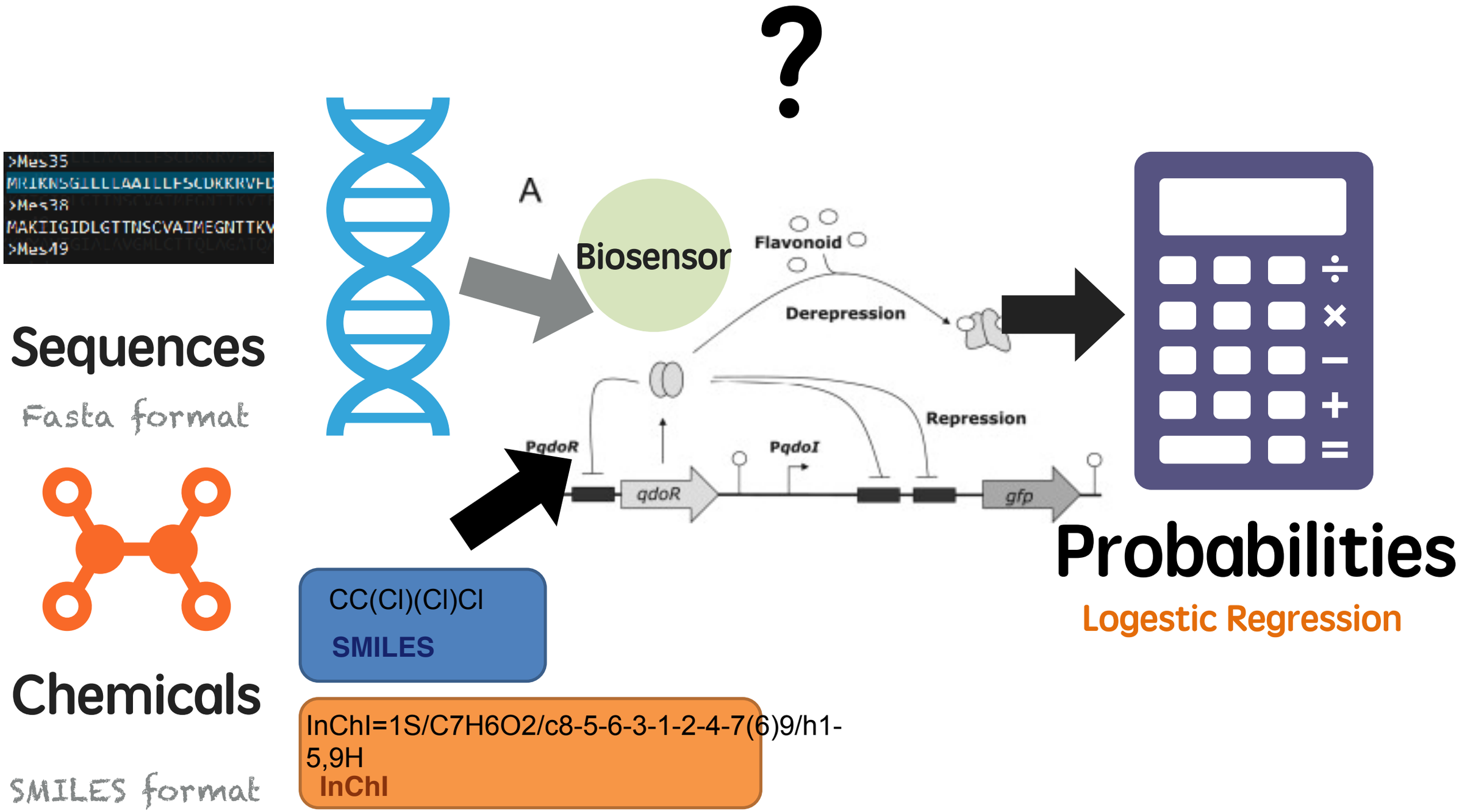
Accuracy Comparison

Estimating the aqueous solubility	General Solubility Equation	Multiple Linear Regression	Our End-to-end Neural Network
Input Features	LogP and melting point (Tm)	LogPoctanol, molecular weight, proportion of heavy atoms in aromatic systems, and number of rotatable bonds and etc. (5-8 of their measured values)	SMILES format chemicals
Training Set	2874 compounds in “Small”, “Medium” and “Large” size	2874 compounds in “Small”, “Medium” and “Large” size	1144 compounds (“Small” – “Large”) in 3-folds cross validation
R-square	0.69	0.67	0.64

- Uversal Approximation Theorem

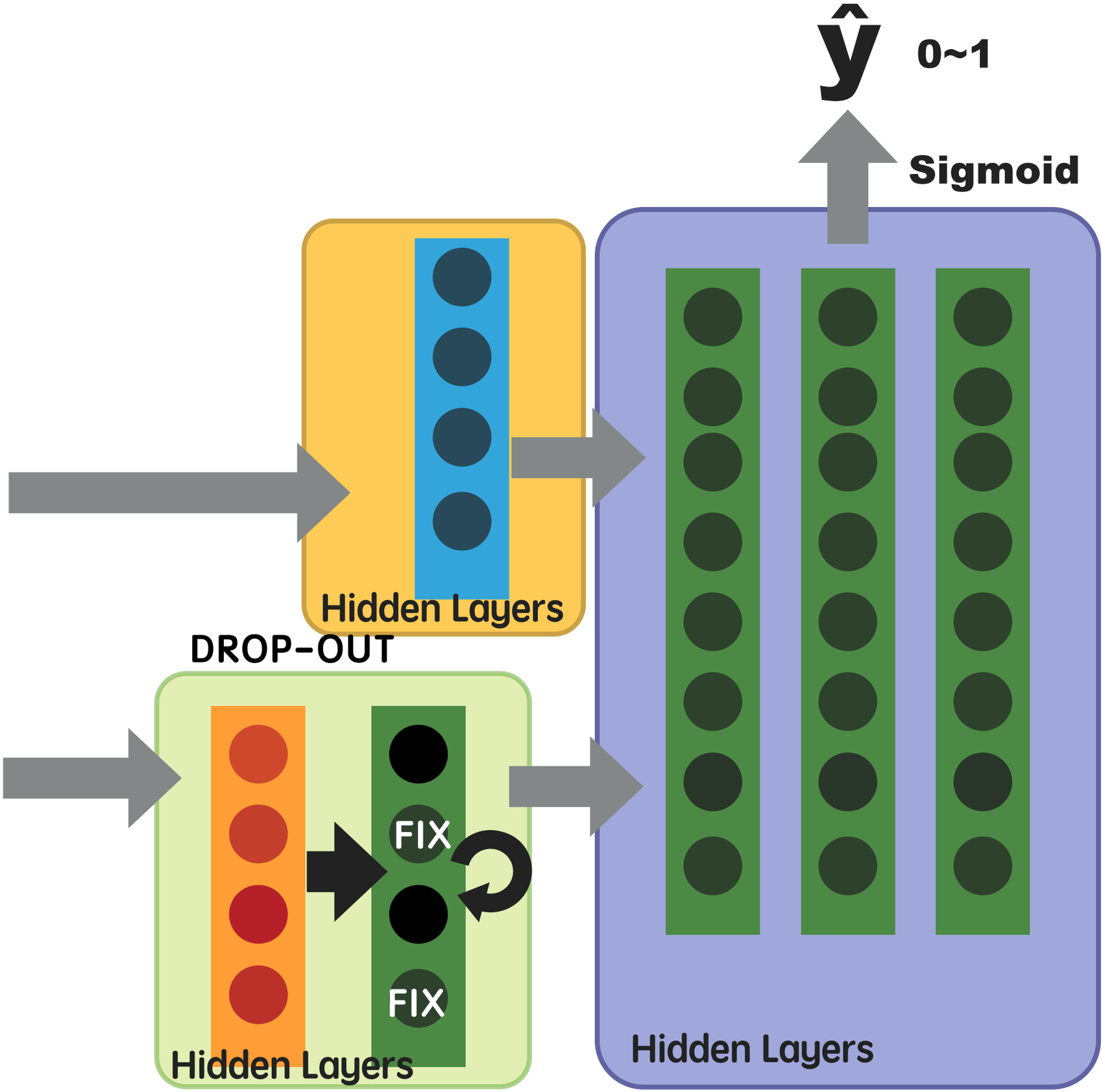
Delaney, J. S. (2004). ESOL: estimating aqueous solubility directly from molecular structure. *Journal of chemical information and computer sciences*, *44*(3), 1000-1005.

Biosensor prediction



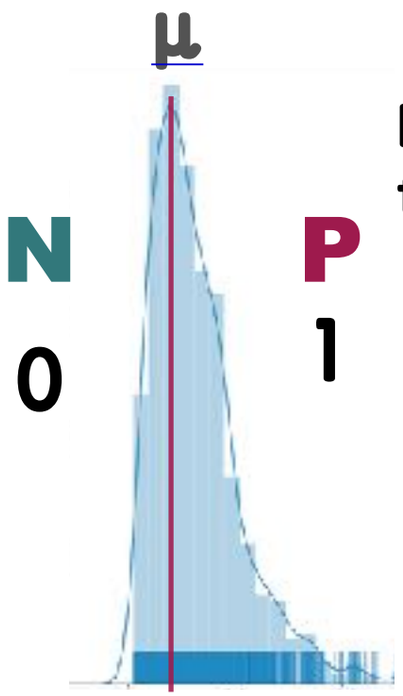
THANK YOU

Ran Duan
University of Manchester
MSc Bioinformatics and Systems Biology
ran.duan-5@postgrad.manchester.ac.uk


ŷ
0~1
Sigmoid
Hidden Layers
DROP-OUT
FIX
FIX
Hidden Layers
Hidden Layers

Functional API

Accuracy (3-folds cross validation)

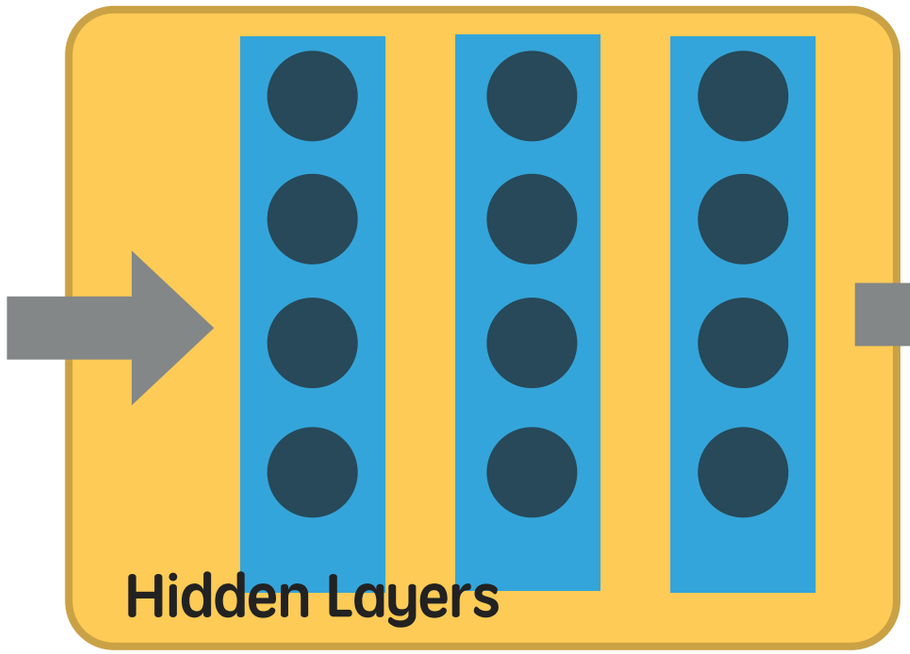


SOLUBILITY

Hardmax the labels

X Y

Compound	Solu
1,1,1,2-Tetrachloroethane	1.0
1,1,1-Trichloroethane	0
1,1,2,2-Tetrachloroethane	1.0



~78%

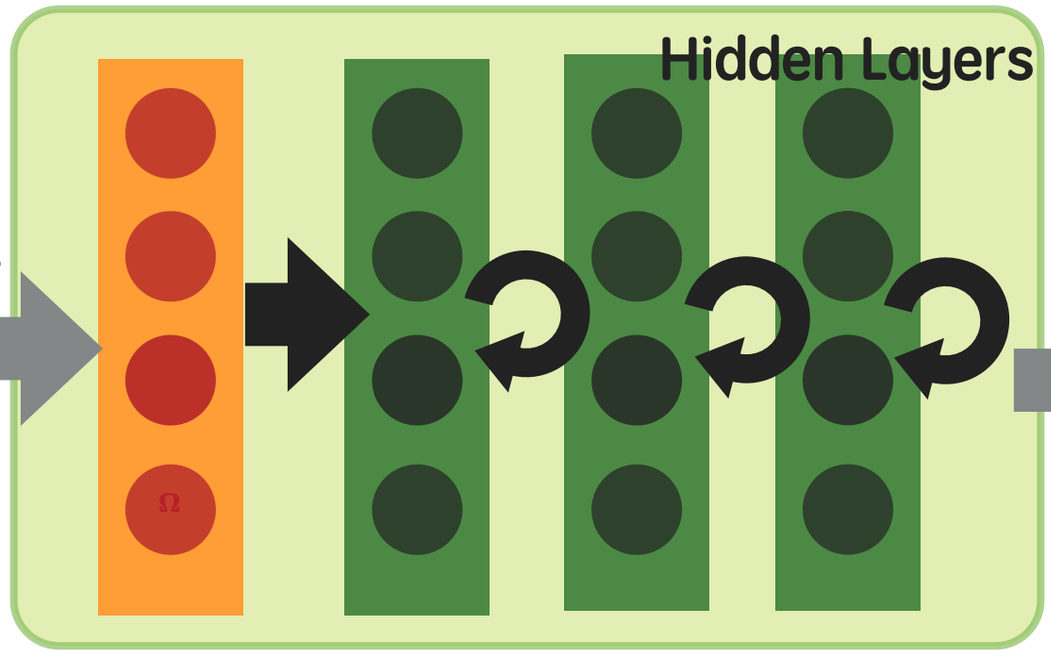
$\hat{y}$

THERMOSTABILITY Y

AMINO ACIDS SEQUENCES

X

Hardmax the labels



$\hat{y}$