

Protein-RNA/ DNA & Deep Learning

Protein-RNA/ DNA Interaction - Motif Analysis

AI vs Machine learning vs Deep Learning

Deep Neural Network

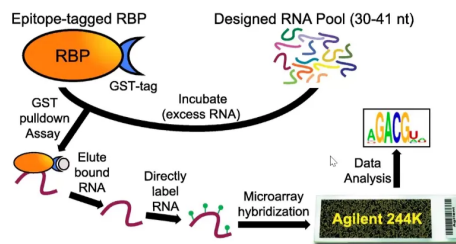


Last Lesson Recaps/ Outline:

- Single-cell sequencing & analysis?
- How is single-cell sequencing done?
- Gene Expression Matrix (Read)?
- Challenges in single-cell data analytics and solutions?
- Visualisation of a large matrix of gene expression data (t-SNE)?

Protein-RNA/ DNA Interaction - Motif Analysis

- **Motif: recurring pattern/ element within the dataset**
 - Molecular biology: the preference pattern for DNA/ RNA/ protein sequence
- **Experiment: find out the binding motif/ pattern**



1. Excess RNA is introduced to tagged protein (Epitope-tagged RBP)
2. Designated RNA will bind with the tagged protein
3. Excess, non-binding RNA is eluted (washed away)
4. Protein is removed from the target RNA
5. RNA is sequenced, aligned and analysed for motif

- **Sequence Alignment to motif**

Table 1: Starting sequences.

#	Sequence
1	AAGAAT
2	ATCATA
3	AAGTAA
4	AACAAA
5	ATTAAA
6	AAGAAT



Table 2: Position Count Matrix.

Position	1	2	3	4	5	6
A	6	4	0	5	5	4
C	0	0	2	0	0	0
G	0	0	3	0	0	0
T	0	2	1	1	1	2

Table 3: Position Probability Matrix.

Position	1	2	3	4	5	6
A	1.00	0.67	0.00	0.83	0.83	0.66
C	0.00	0.00	0.33	0.00	0.00	0.00
G	0.00	0.00	0.50	0.00	0.00	0.00
T	0.00	0.33	0.17	0.17	0.17	0.33

1. Alignment of sequence (Table 1: First base is always A; Last base is either A or T)
2. Position Counting (Table 2)

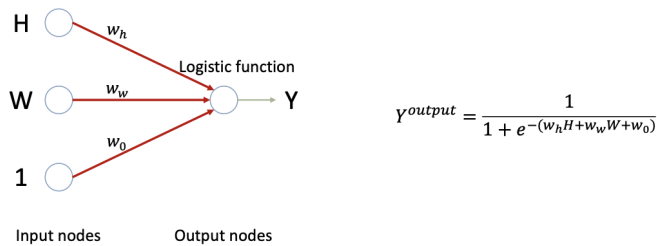
3. Turn the Position Count Matrix into the Probability Matrix (Table 3)

AI vs Machine learning vs Deep Learning

- **Why do we care about health data?**
 - For accurate diagnosis
 - Needed to train/ deep learning for more accurate disease screening
- **AI vs Machine learning vs Deep Learning**
 - **Artificial Intelligence(AI):** Any techniques that mimic human behaviour
eg. Robot (fixed instructions)
 - **Machine Learning (ML):** A subset of AI; perform a specific task without instructions; relying on patterns from the data
eg. Self-driving car (No explicit instruction)
 - **Deep Learning (DL):** A subset algorithms of ML; multi-layer neural networks

Deep Neural Network

- **Problem of Logistic regression**



- Relationships among variables are much more complicated/ nonlinear
- **Solution: Deep Neural Networks**
 - Increase the number of nodes, layer/ Add non-linear functions
 - Each node is still supported by logistic regression (Output of layer 1 = Input of layer 2)
 - Developing a fully-connected layers