
Machine Learning and Bioinformatics

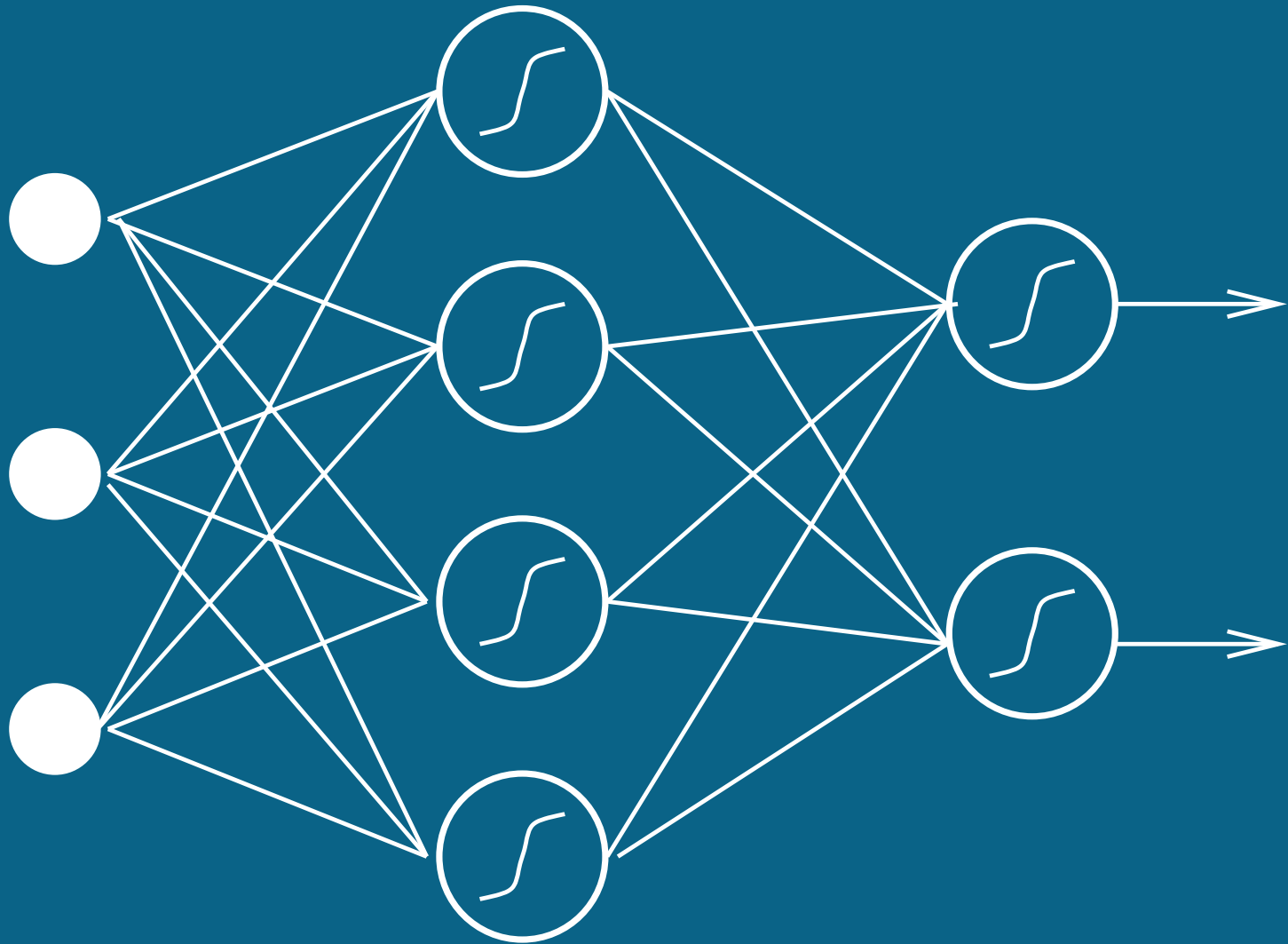
Dirk Husmeier

Biomathematics & Statistics Scotland (BioSS)
JCMB, The King's Buildings, Edinburgh EH9 3JZ
United Kingdom

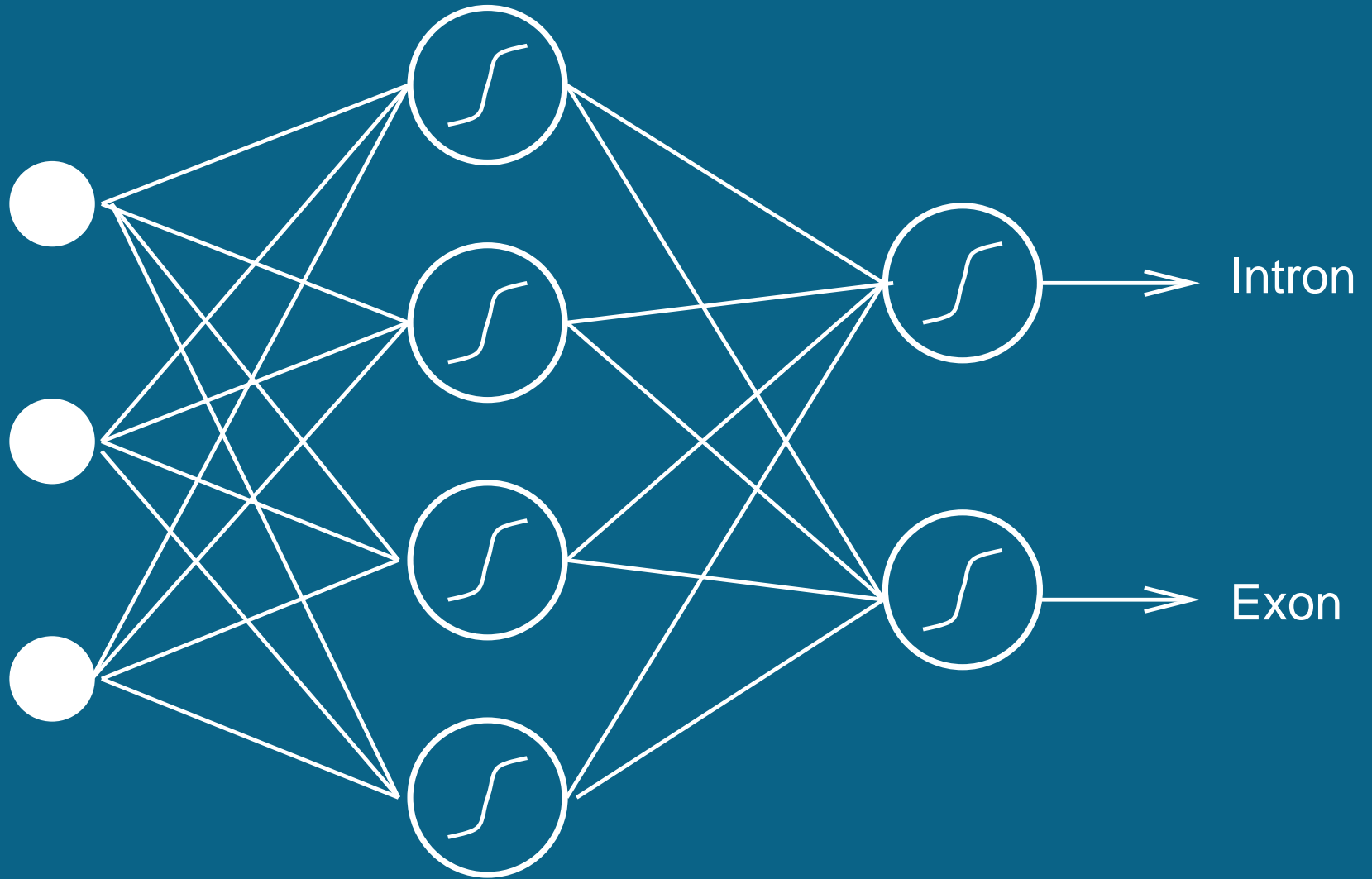
<http://www.bioss.ac.uk/~dirk>

Limitations of Artificial Intelligence (AI)

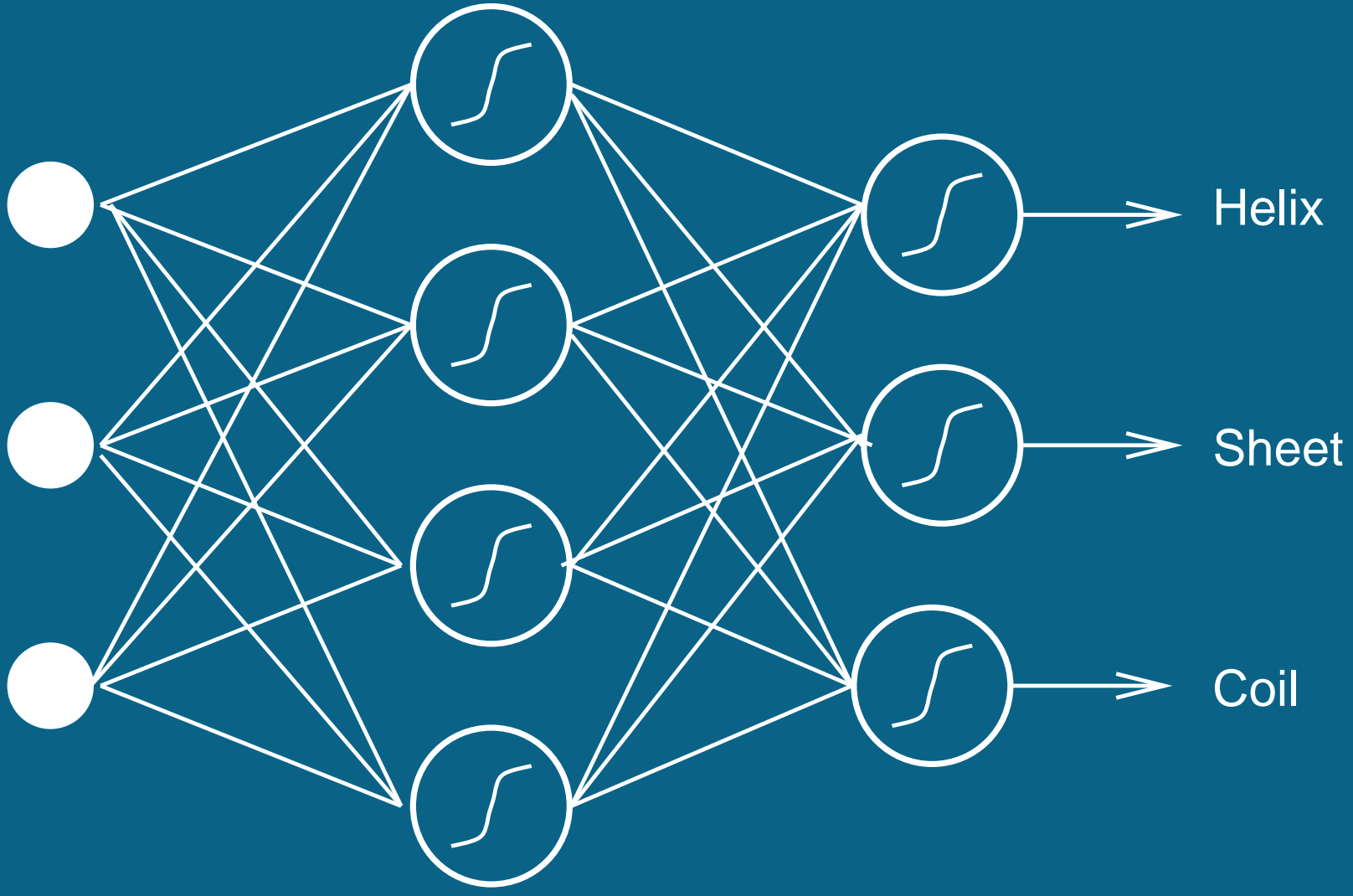
Neural Computation



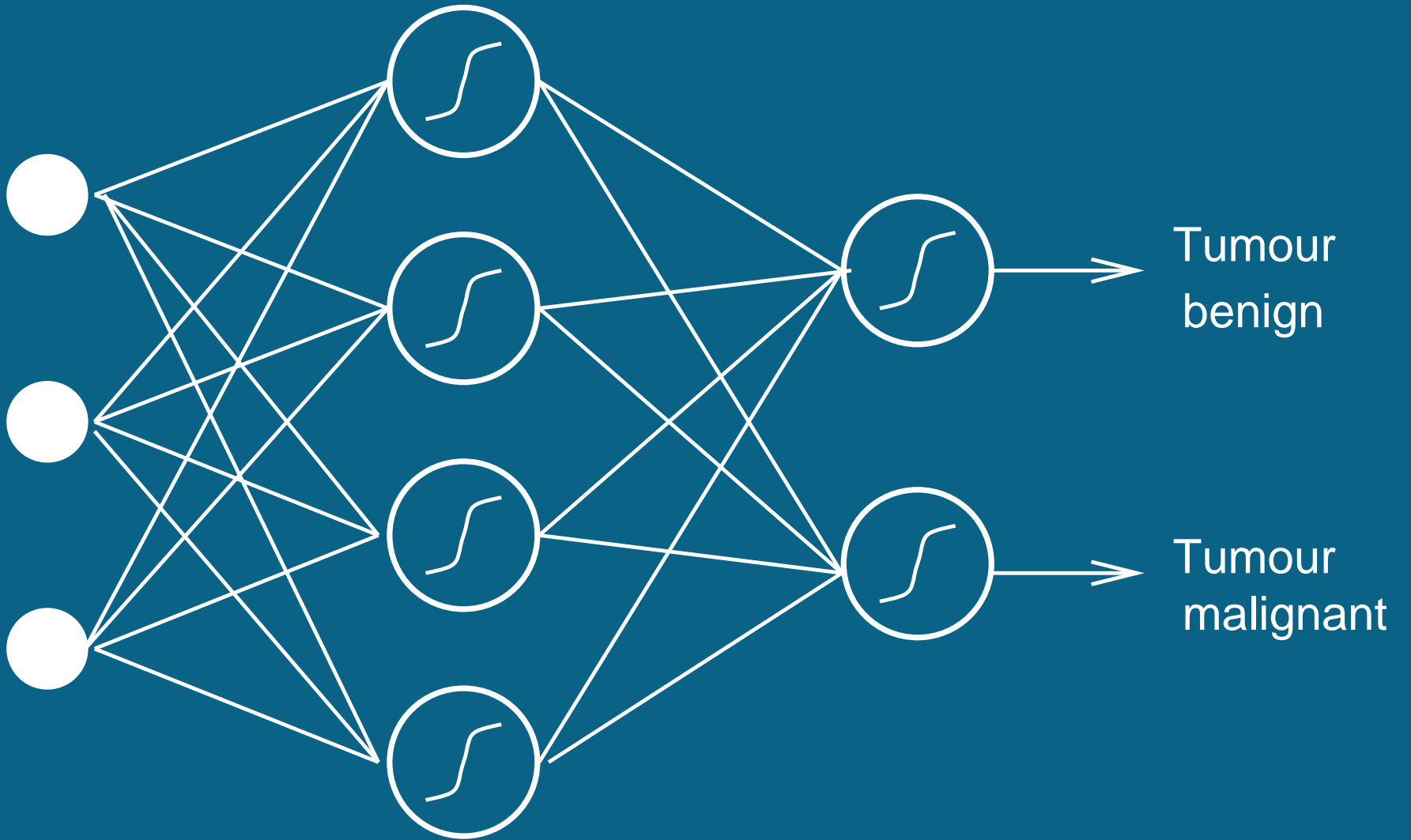
GENOME DATA

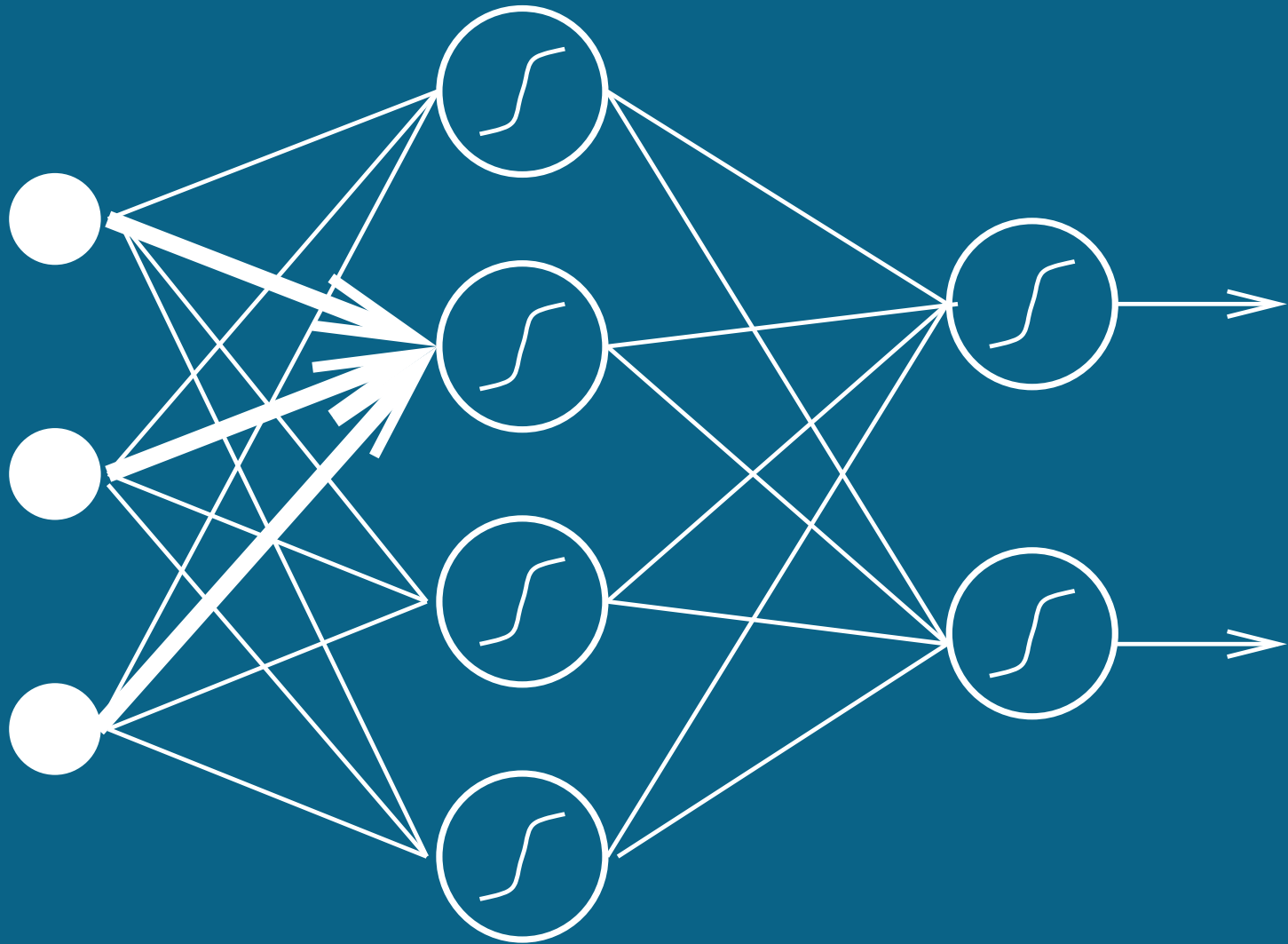


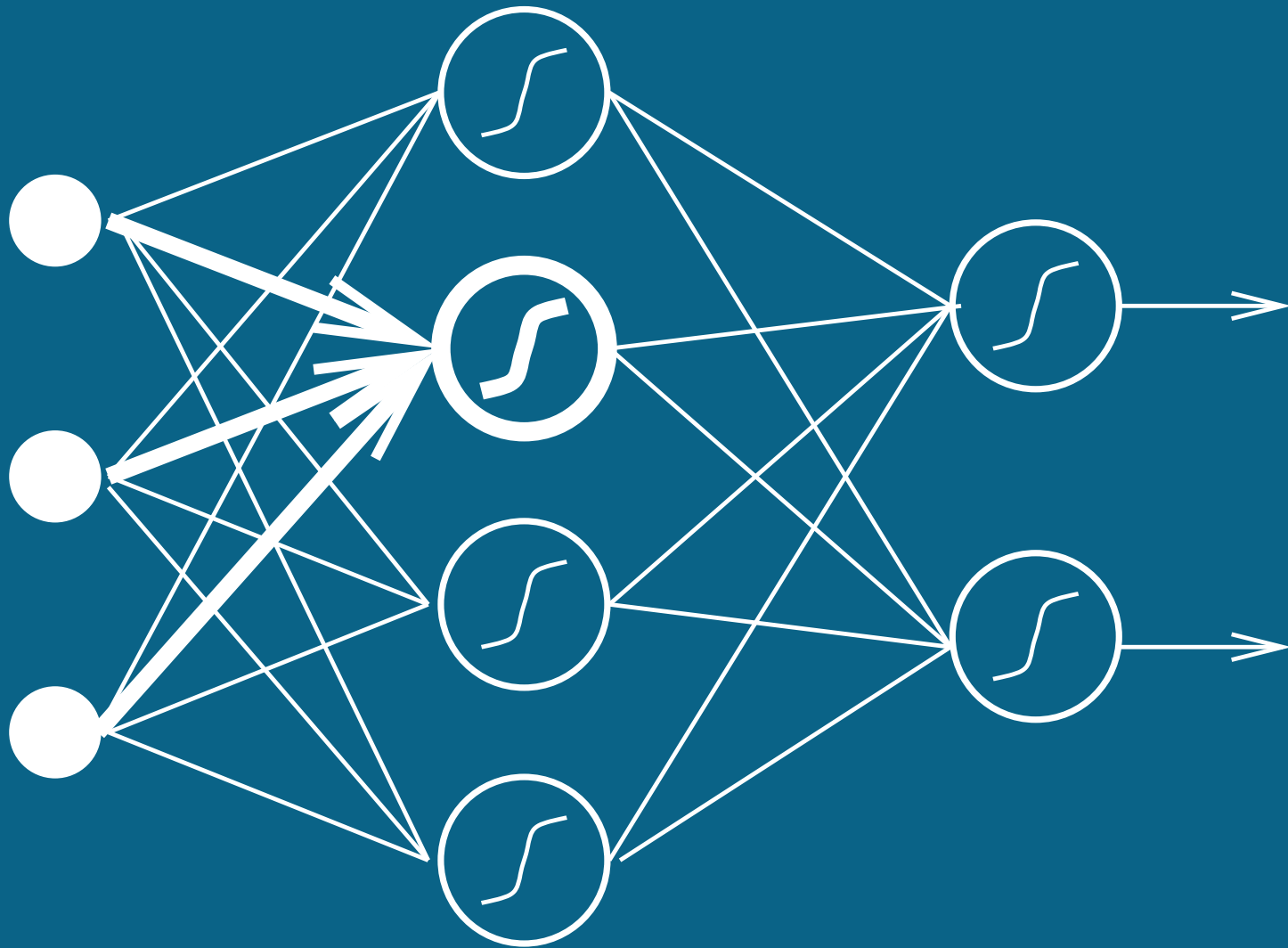
PROPHET Z-MHOTO

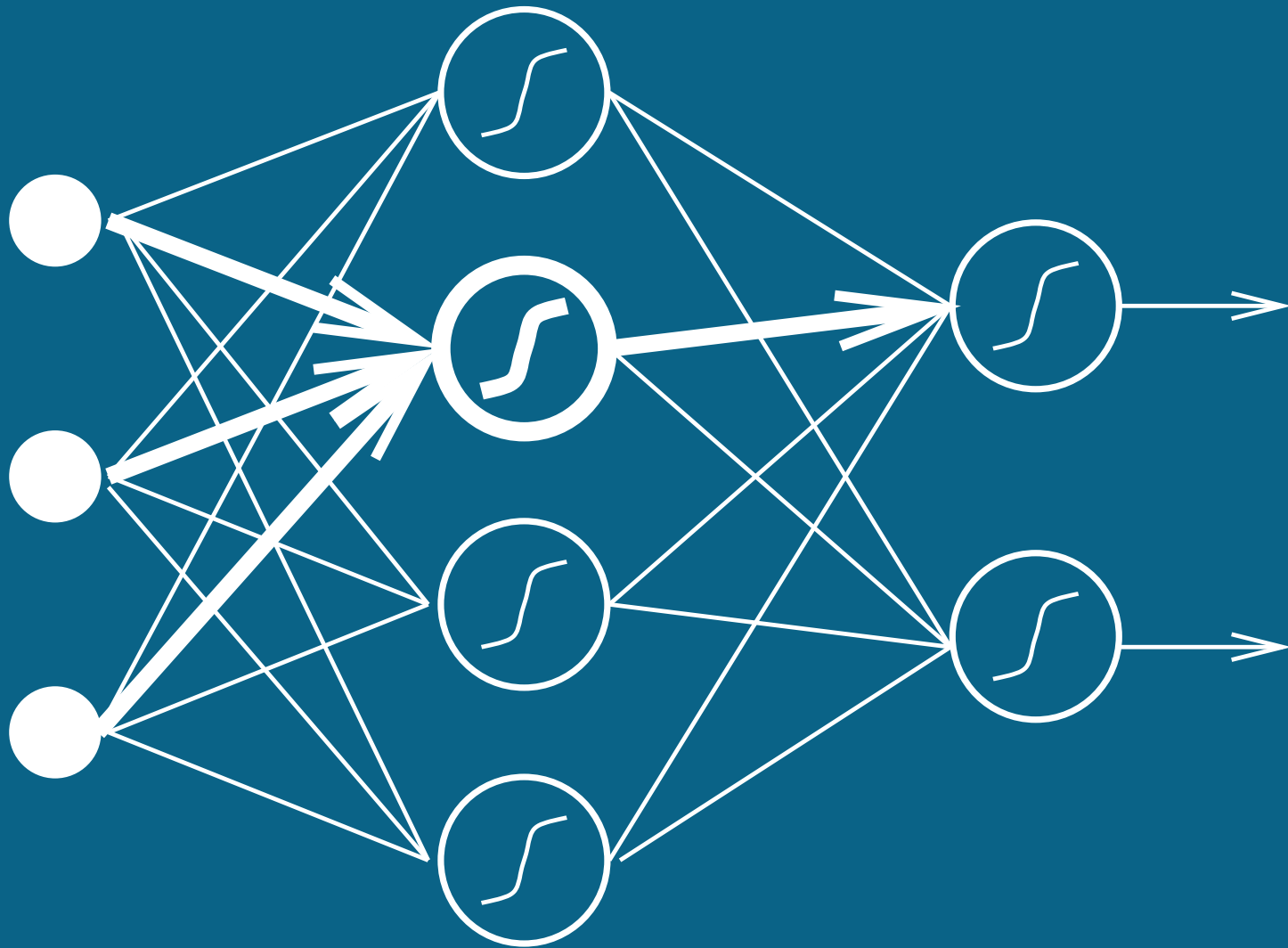


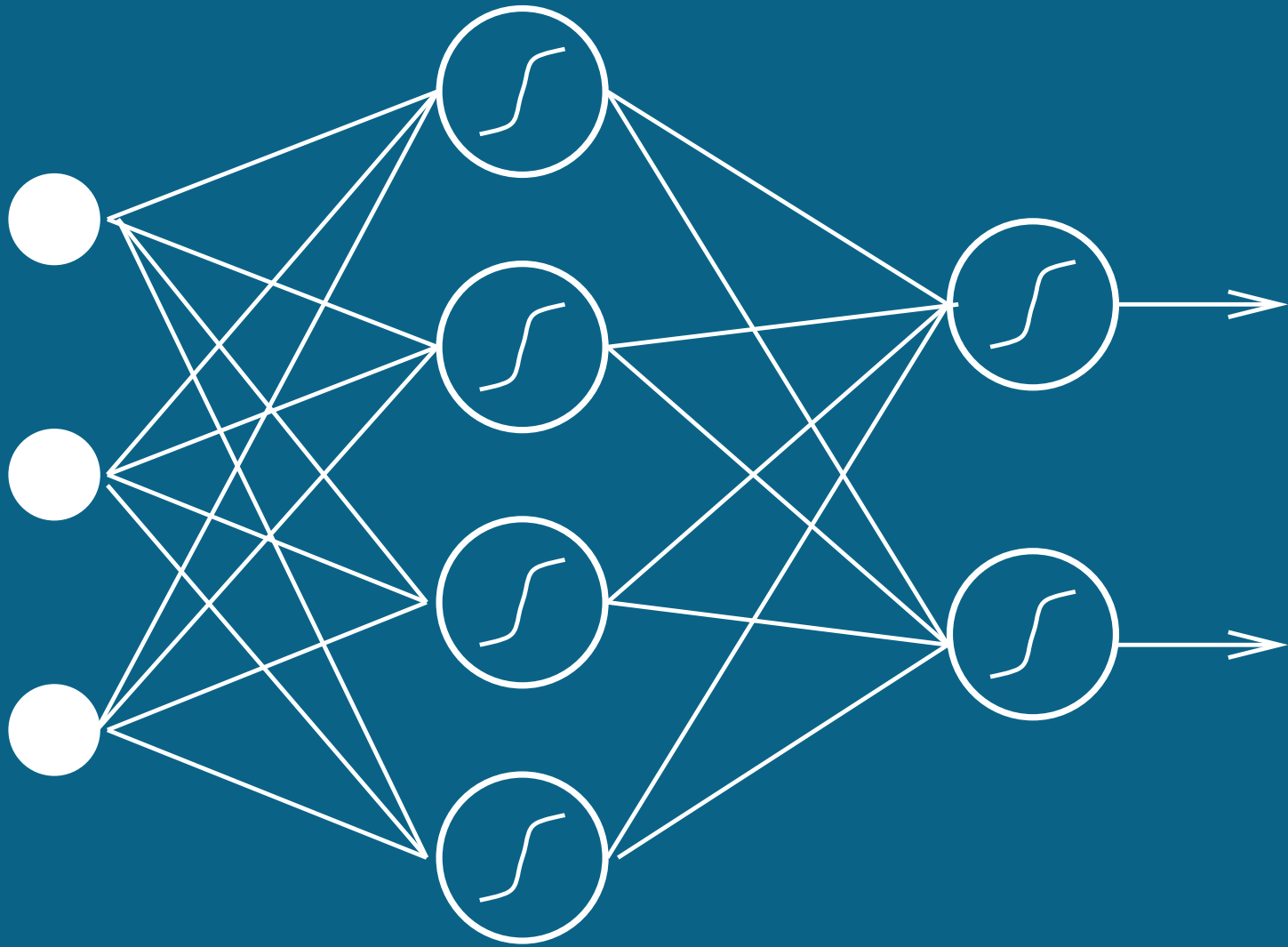
ΥΑΡΡΑΟΡΟΡΟ—M



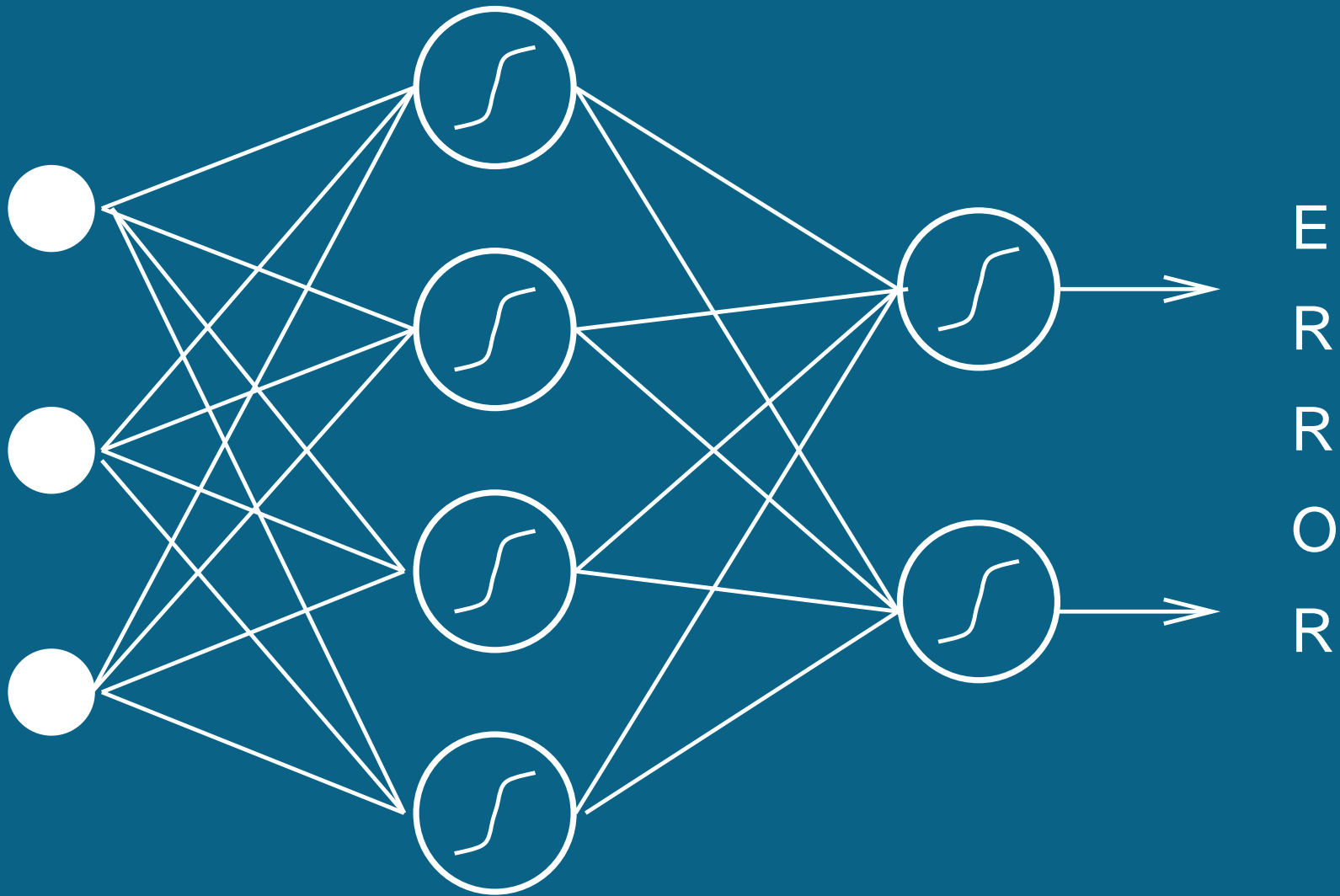


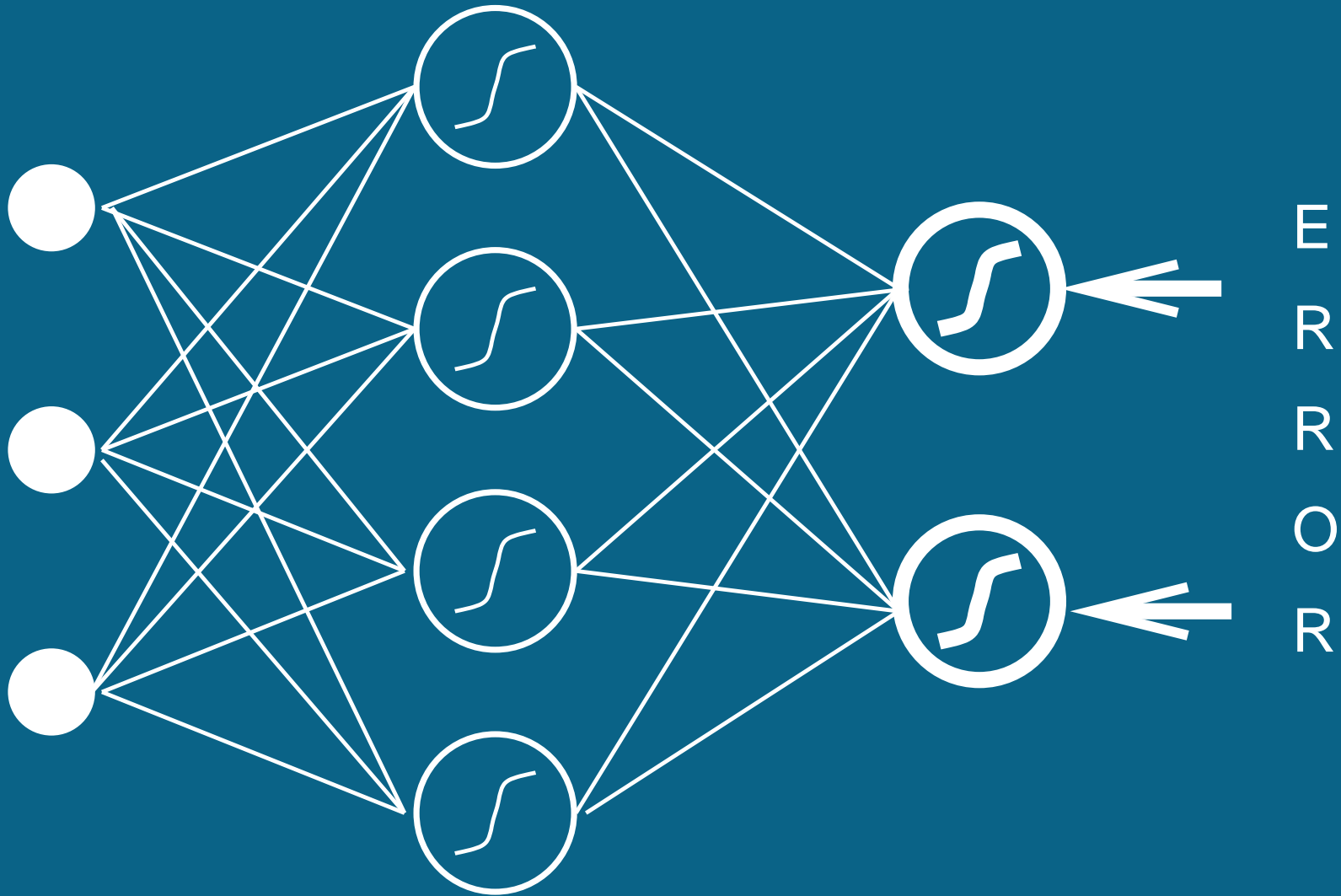


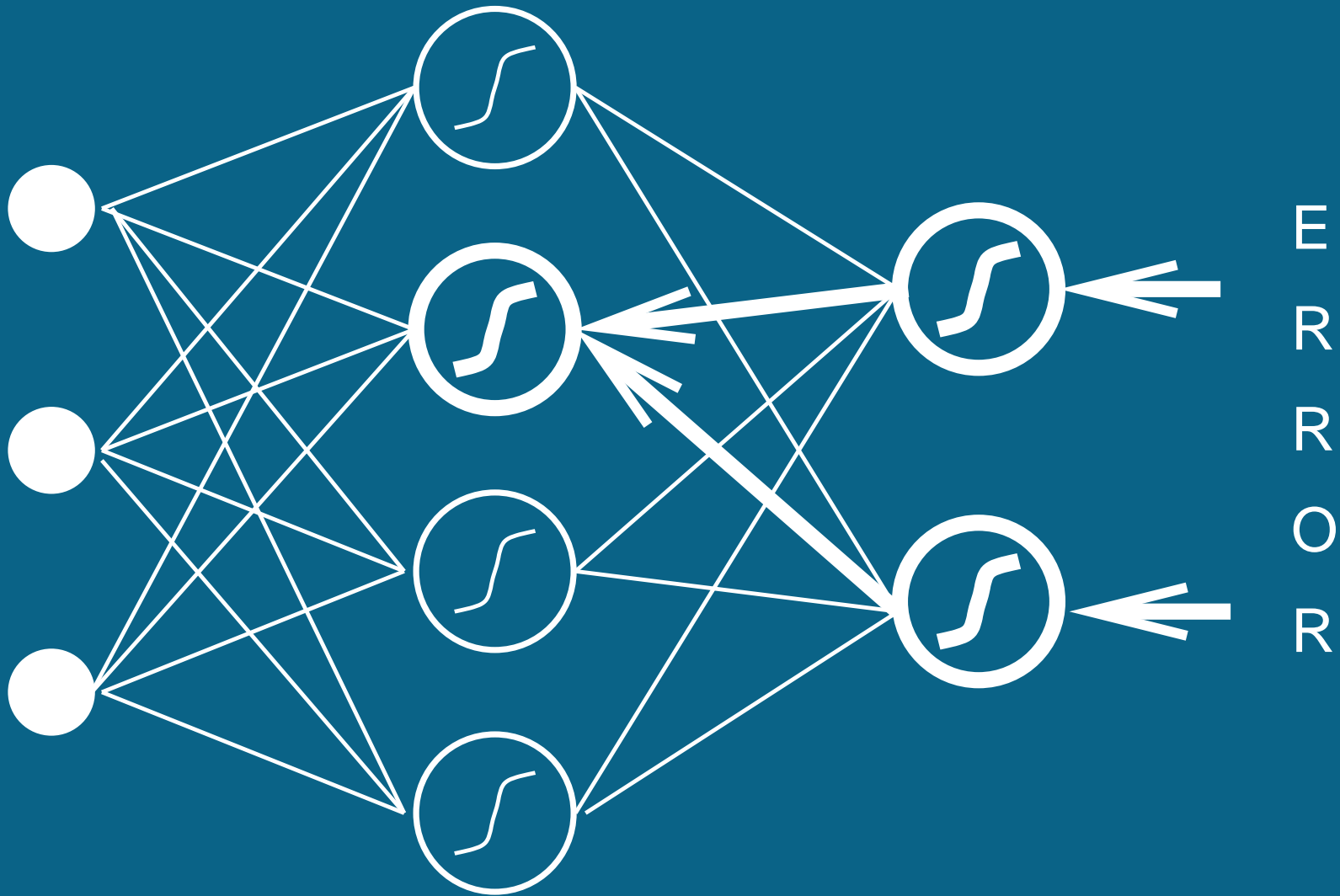


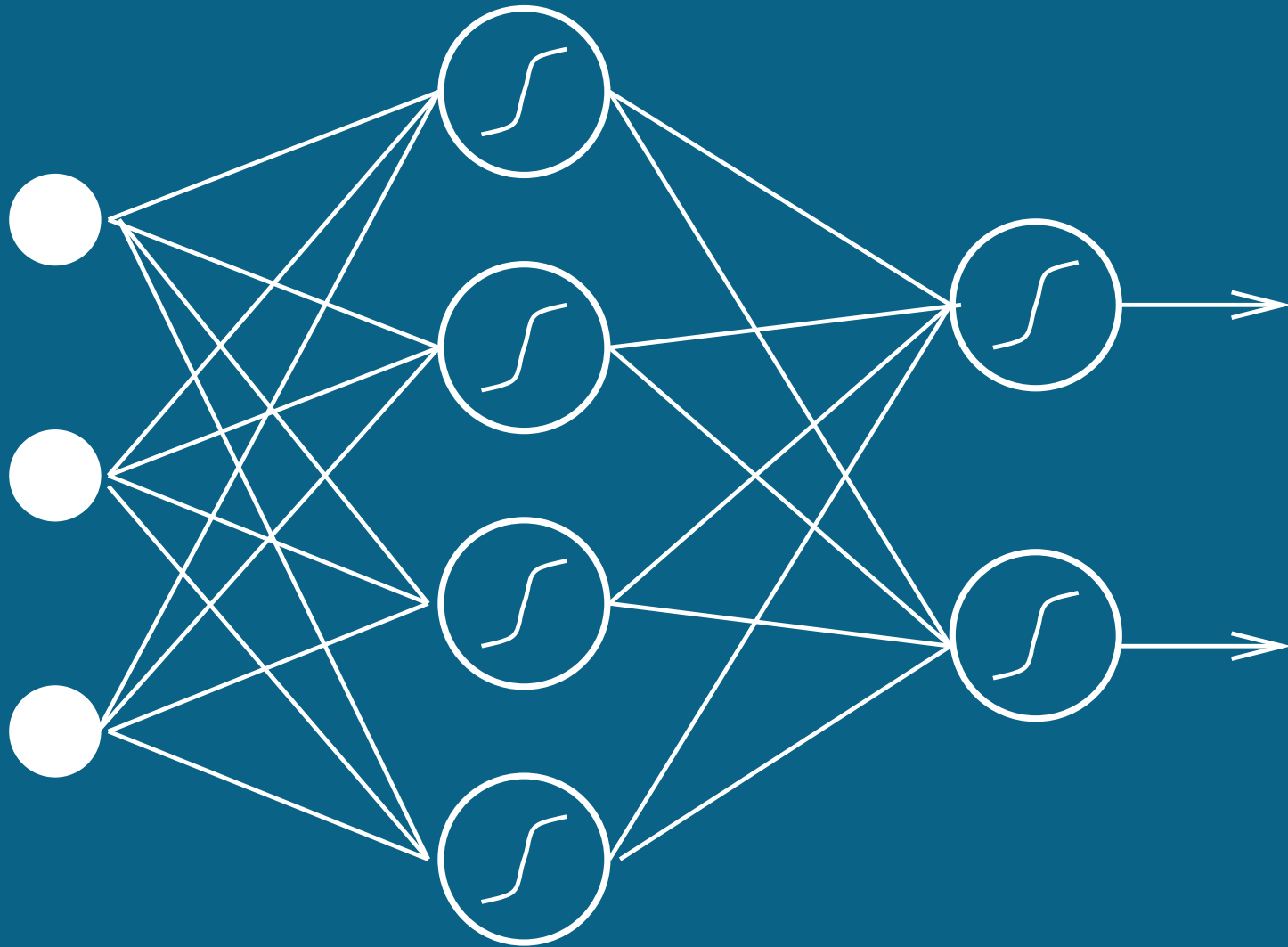


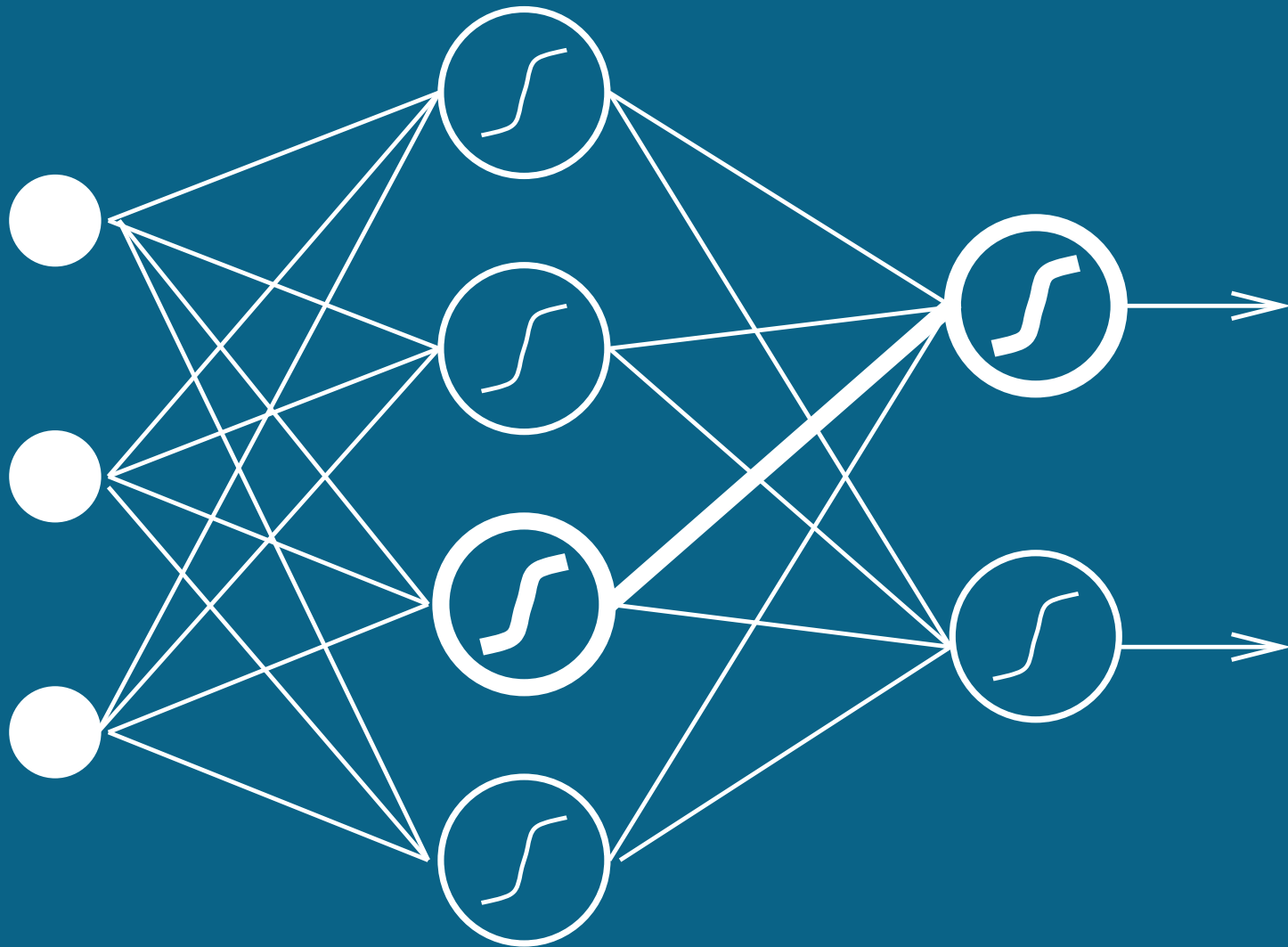
NO--HC--DMRT

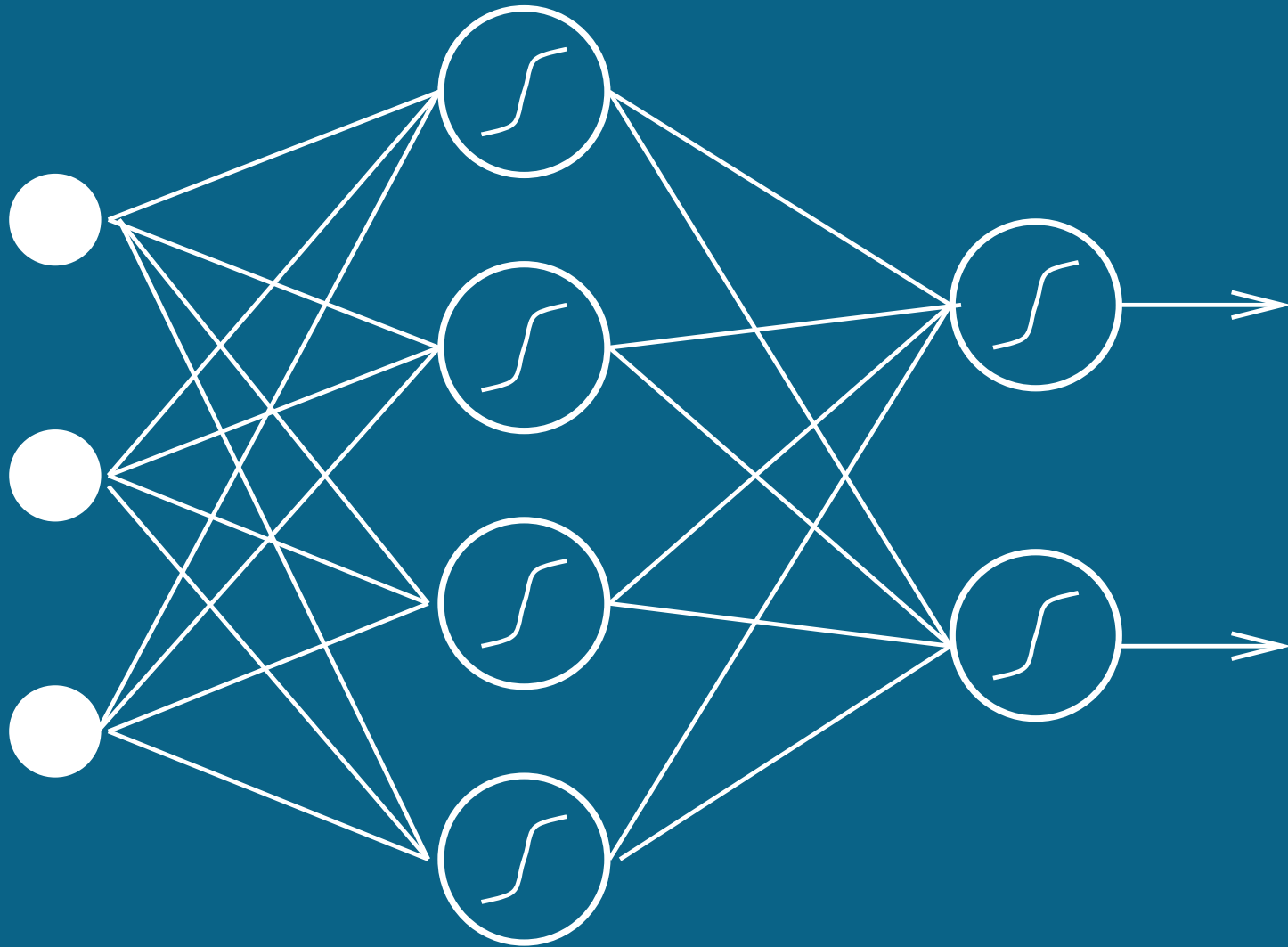


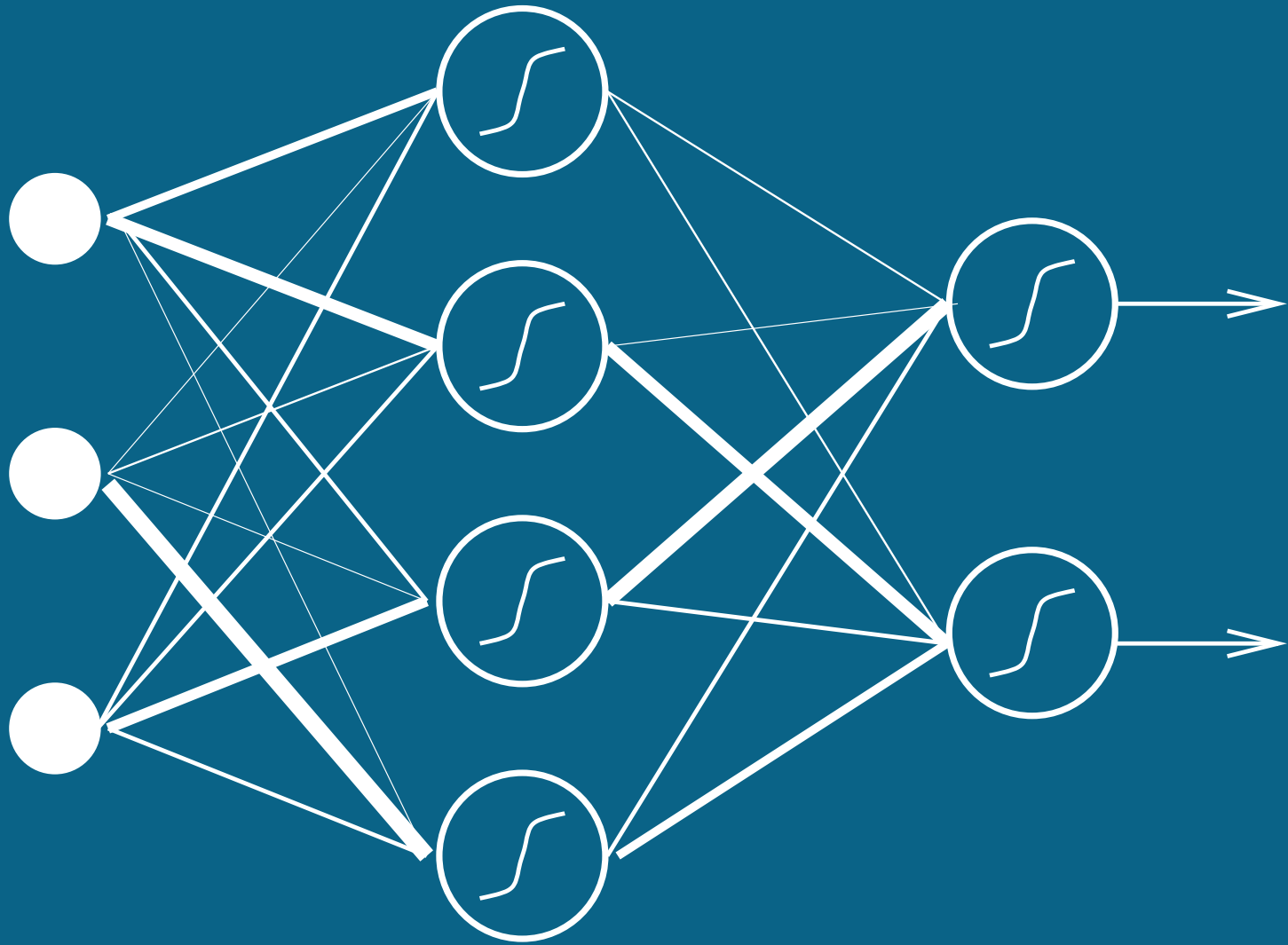




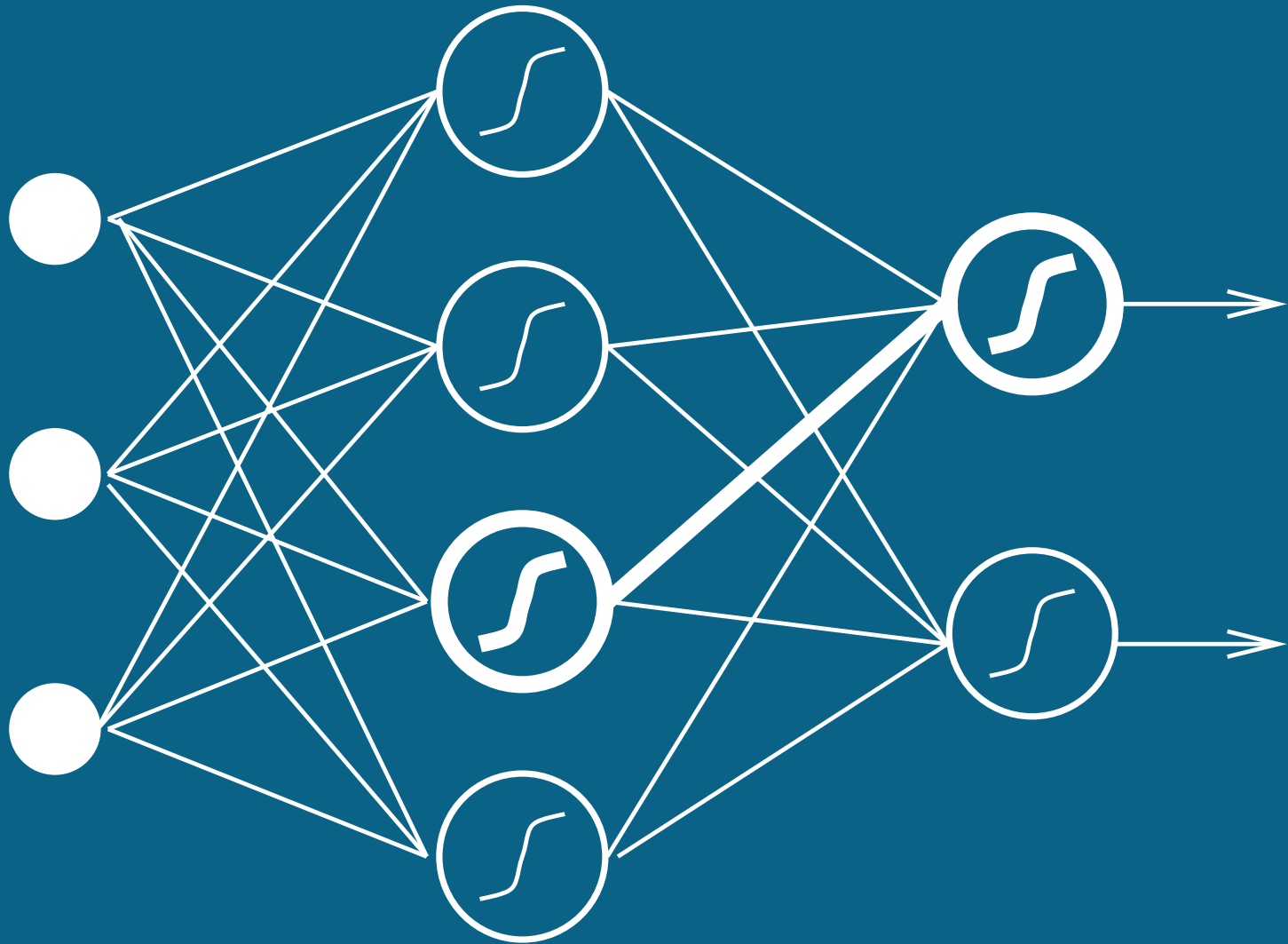


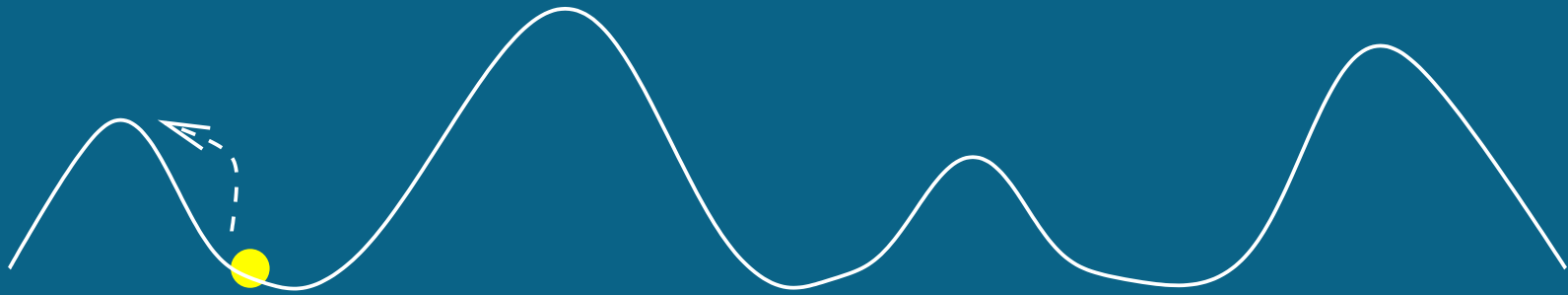


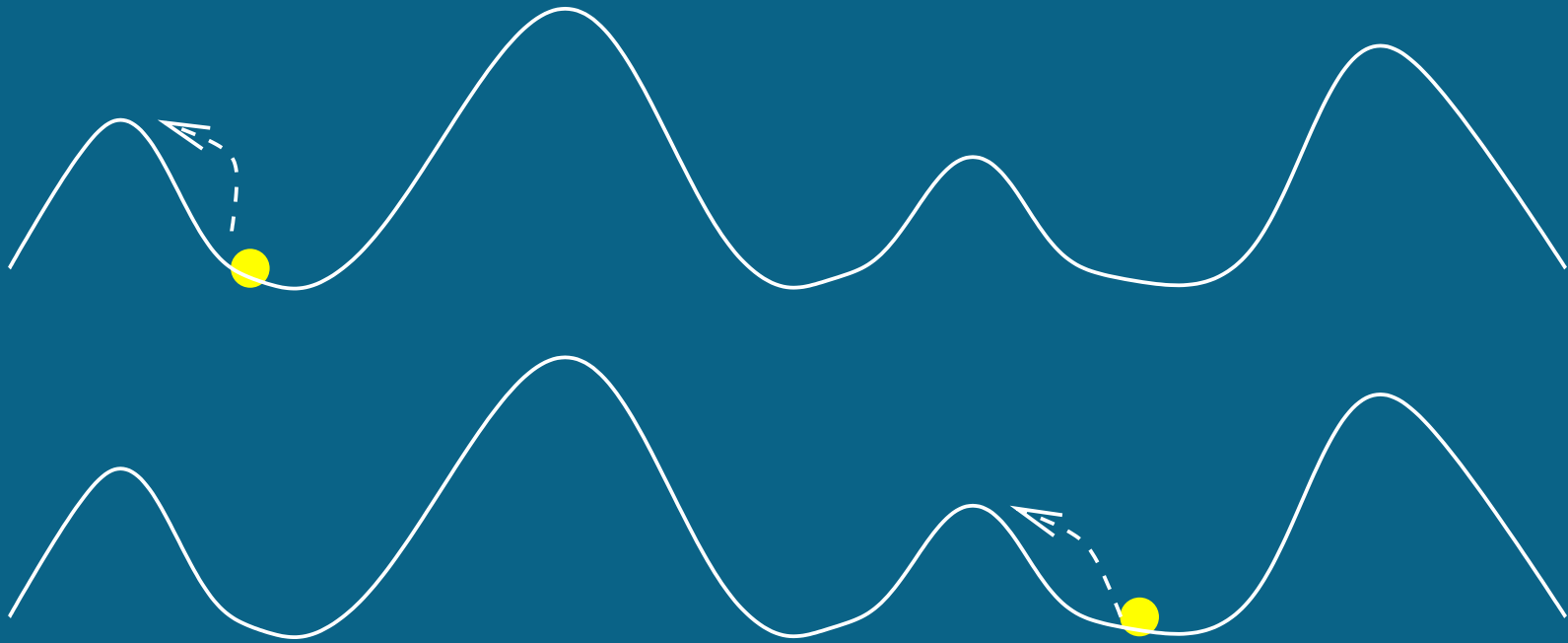


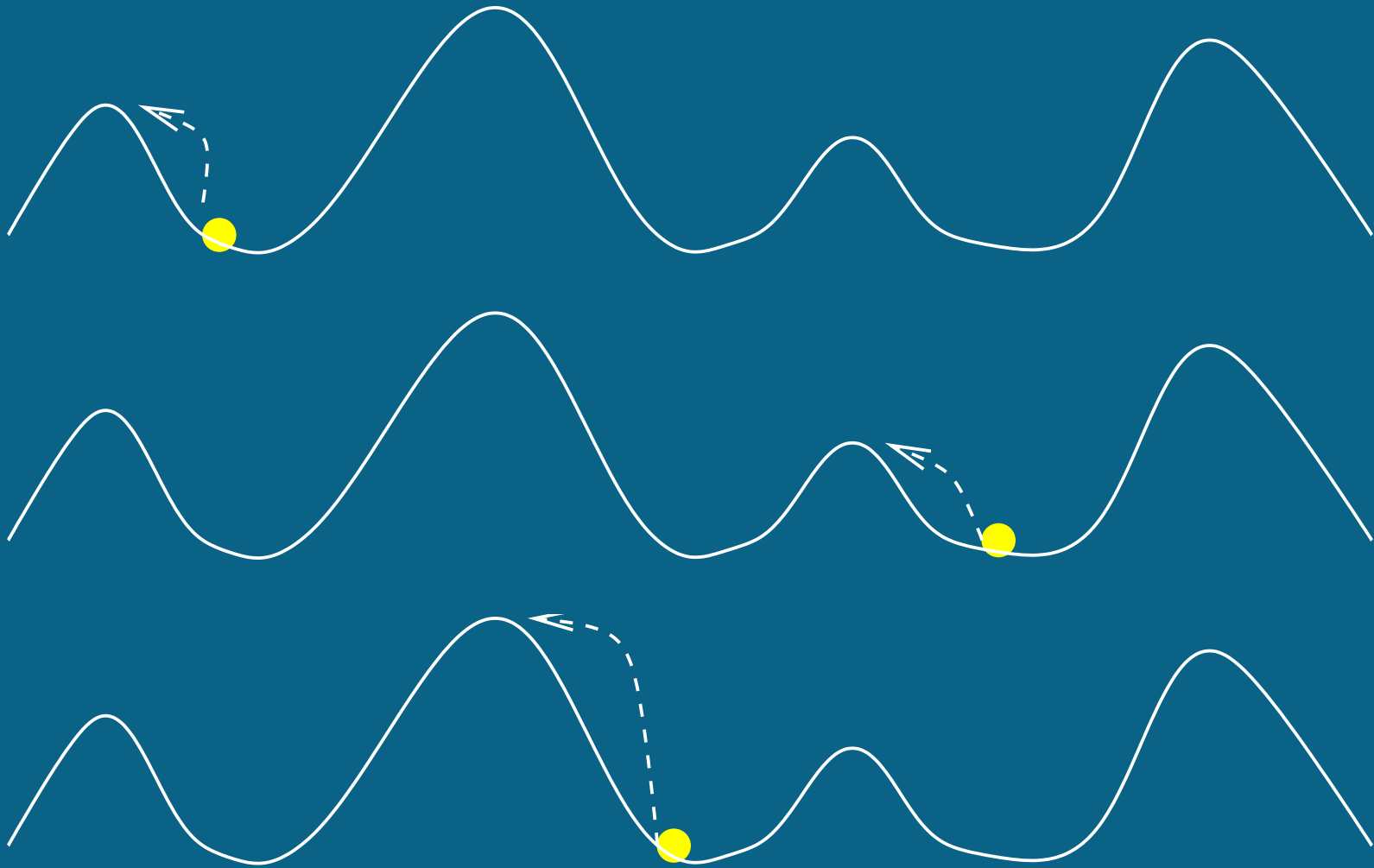


Learning

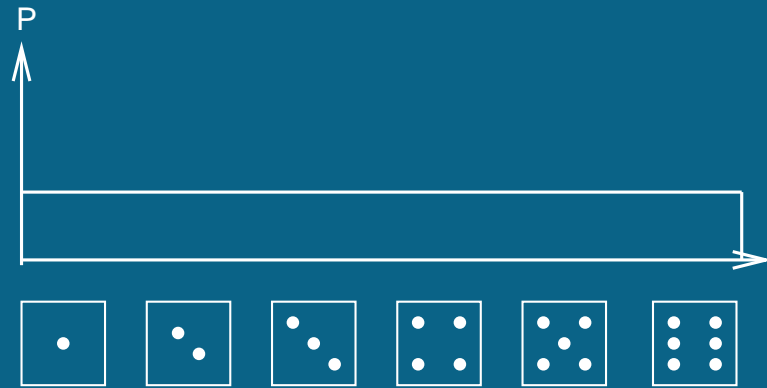
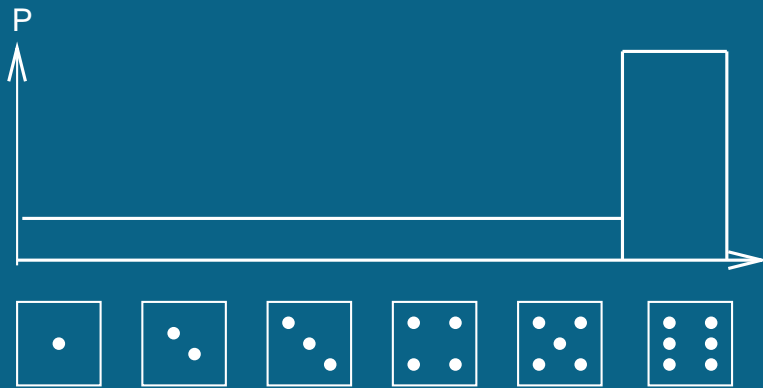
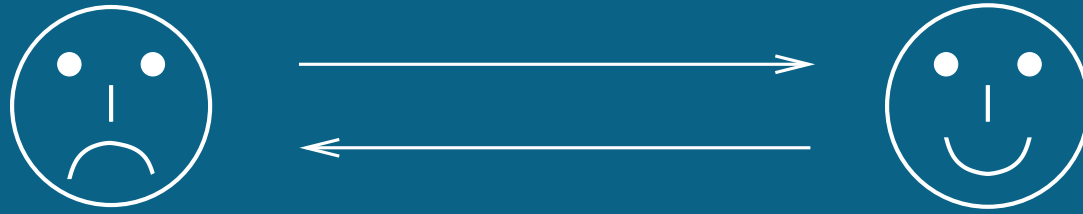




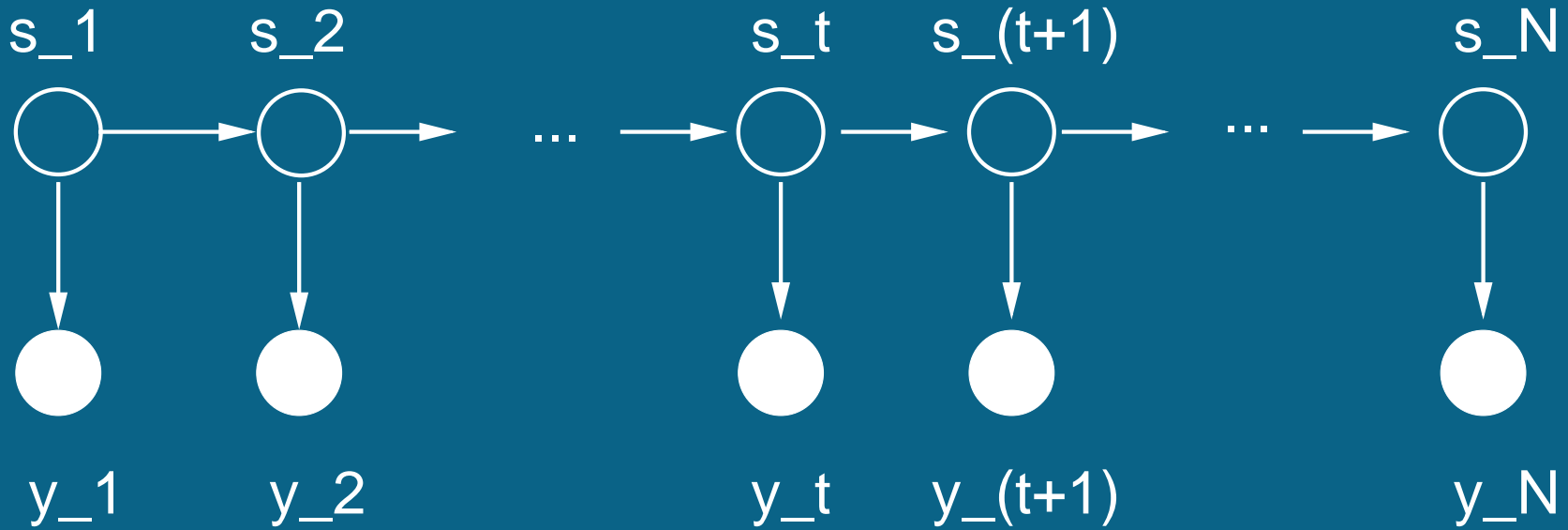




Hidden Markov Models





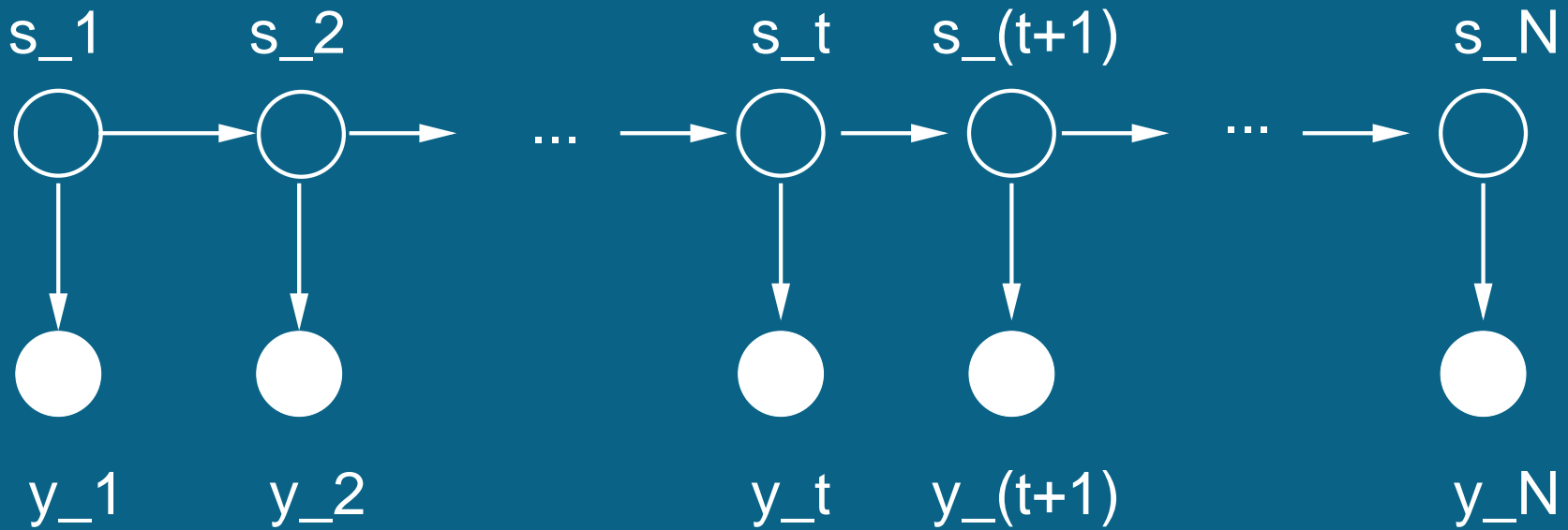




Non-coding

Coding

Non-coding



- Two DNA sequences:
Sequence 1 → ACGTTGCA
Sequence 2 → ACGTGCAAT

- Two DNA sequences:
Sequence 1 → ACGTTGCA
Sequence 2 → ACGTGCAT

- Direct alignment:
A C G T T G C A
A C G T G C A T

- Two DNA sequences: Sequence 1 → ACGTTGCA
Sequence 2 → ACGTGCAT

- Direct alignment:

A	C	G	T	T	G	C	A
A	C	G	T	G	C	A	T

- Alignment with gaps:

A	C	G	T	T	G	C	A	-
A	C	G	T	-	G	C	A	T

- Two DNA sequences:

Sequence 1	→	ACGTTGCA
Sequence 2	→	ACGTGCAT

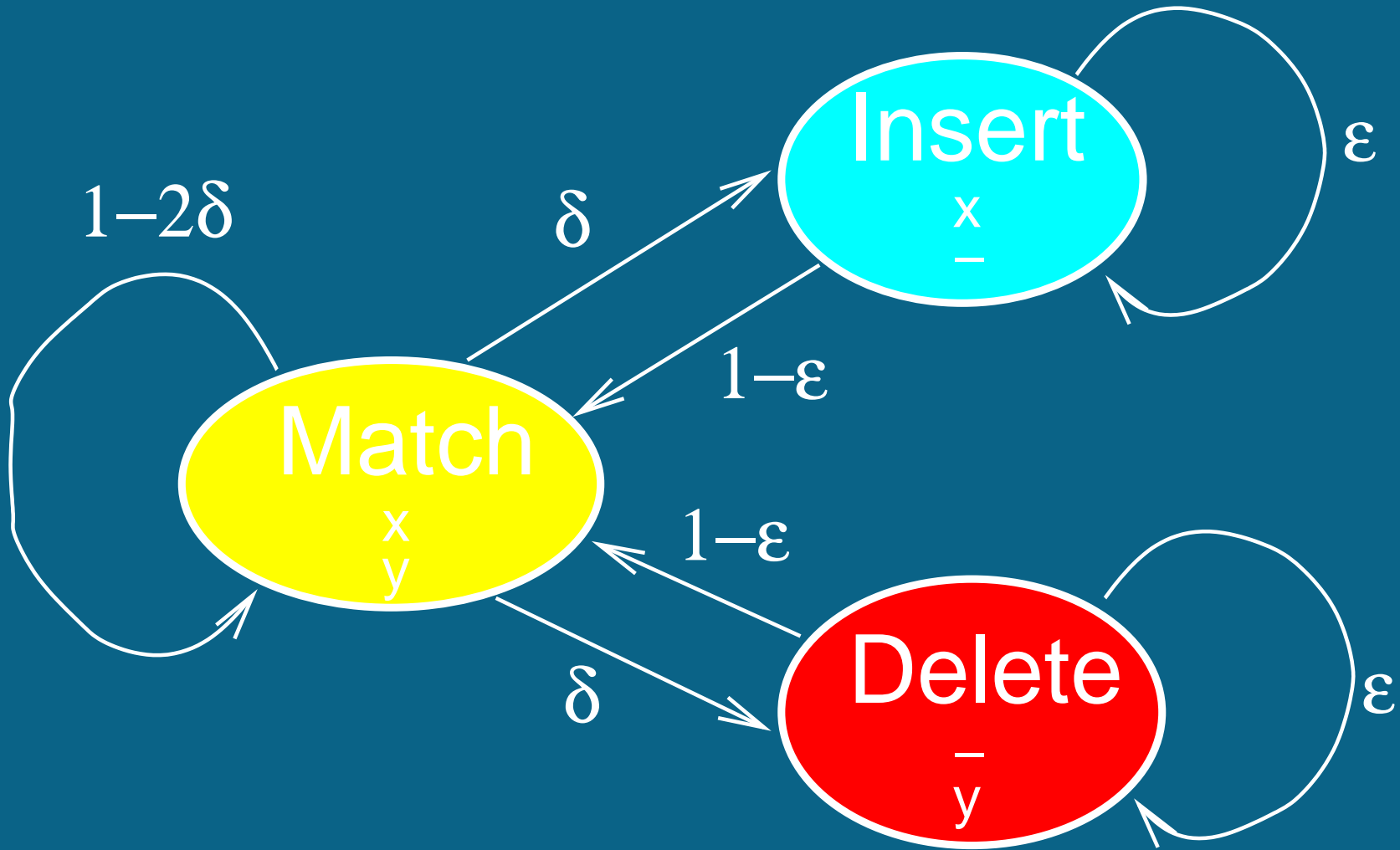
- Direct alignment:

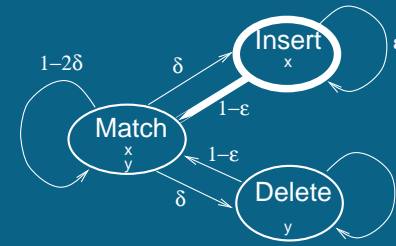
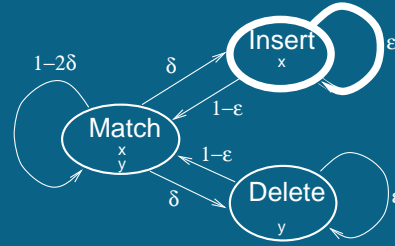
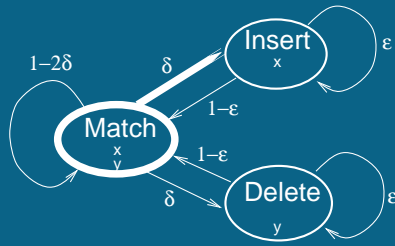
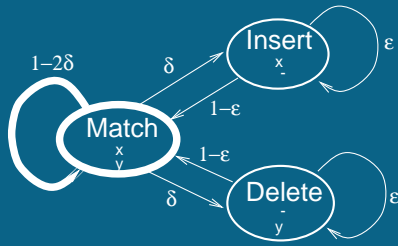
A	C	G	T	T	G	C	A
A	C	G	T	G	C	A	T

- Alignment with gaps:

A	C	G	T	T	G	C	A	-
A	C	G	T	-	G	C	A	T

- Introduce insertions and deletions.





T

C

T

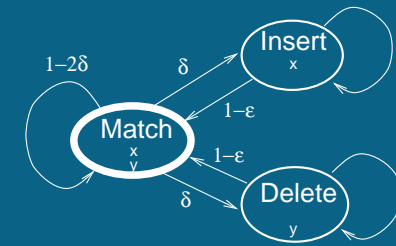
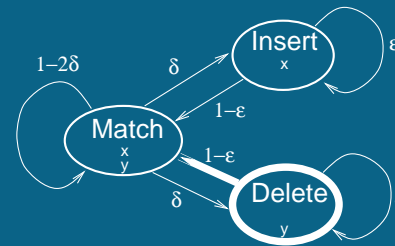
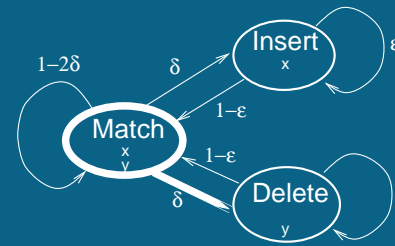
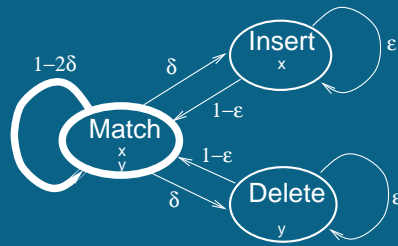
A

T

G

-

-



C

T

-

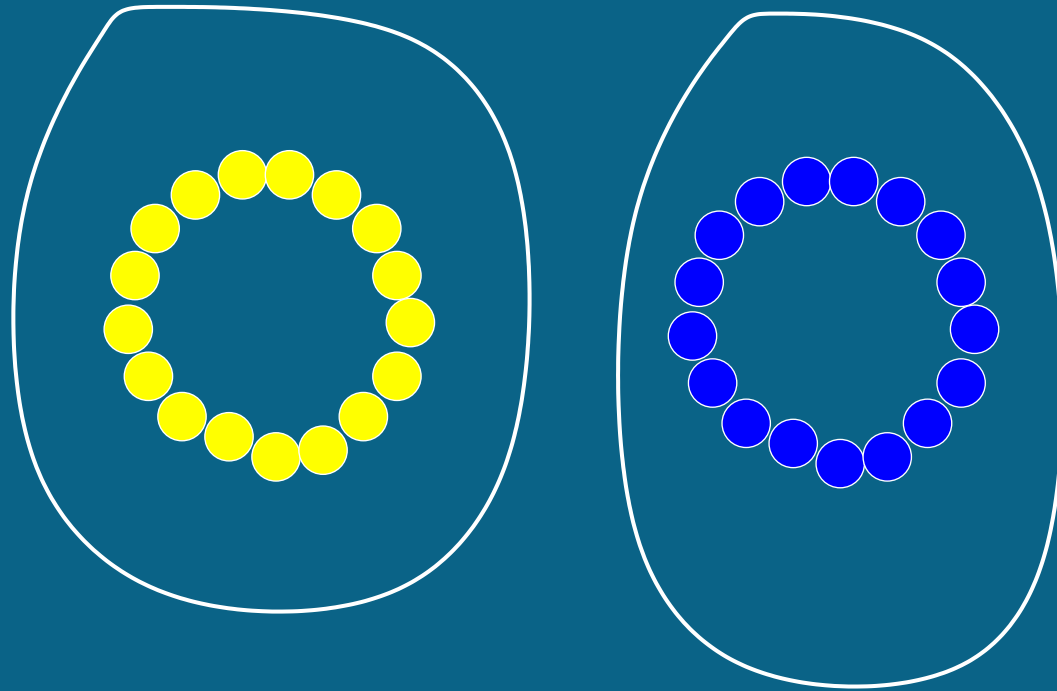
T

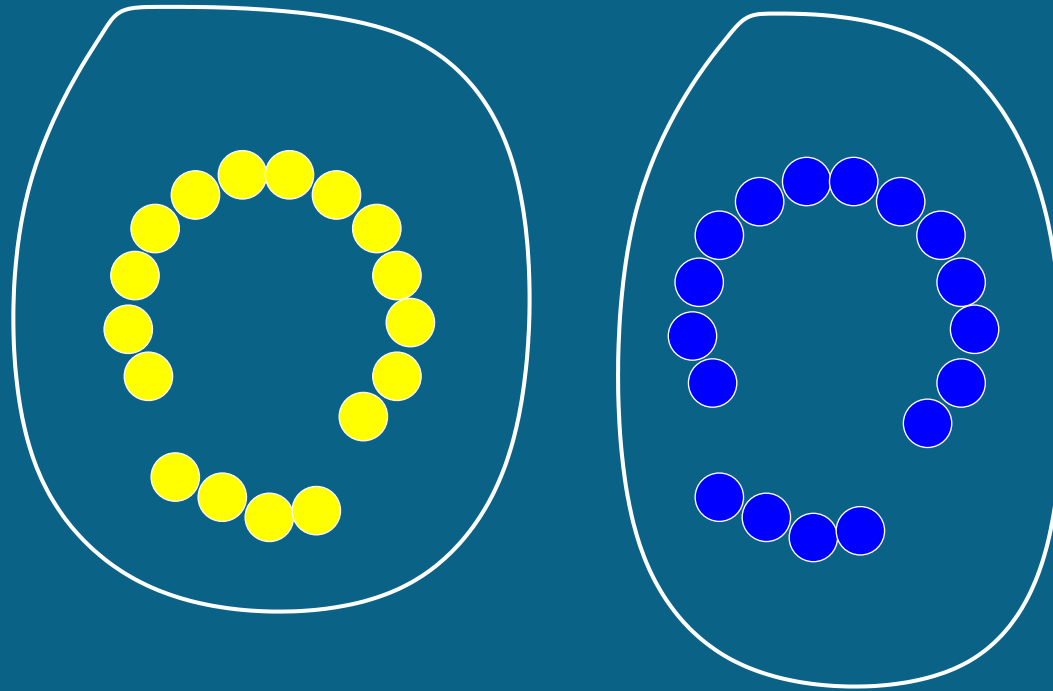
C

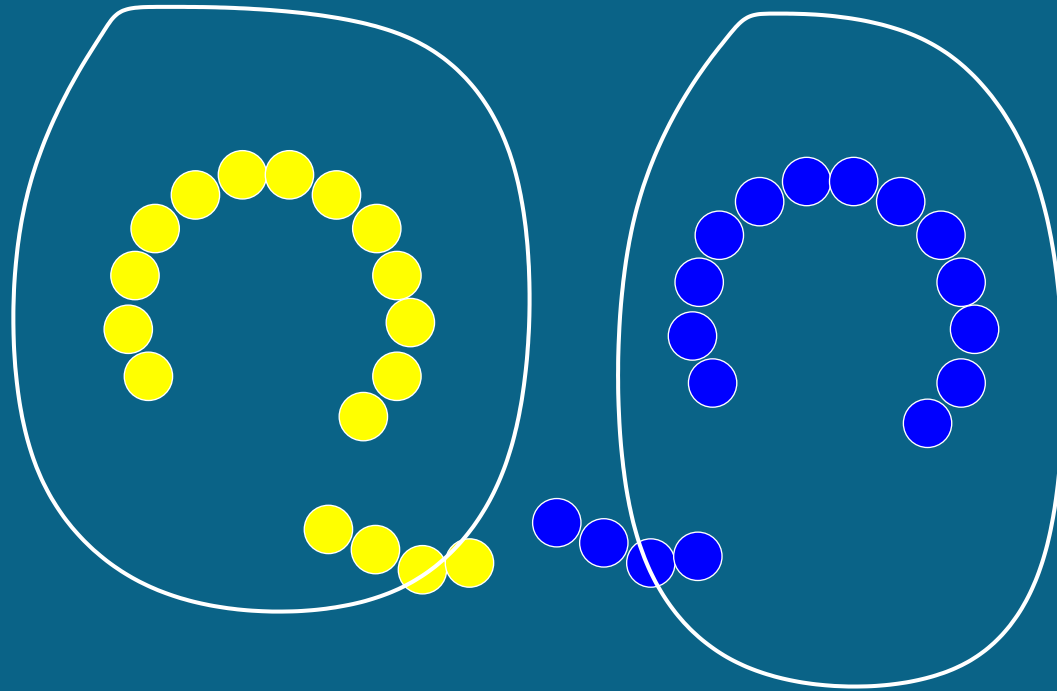
A

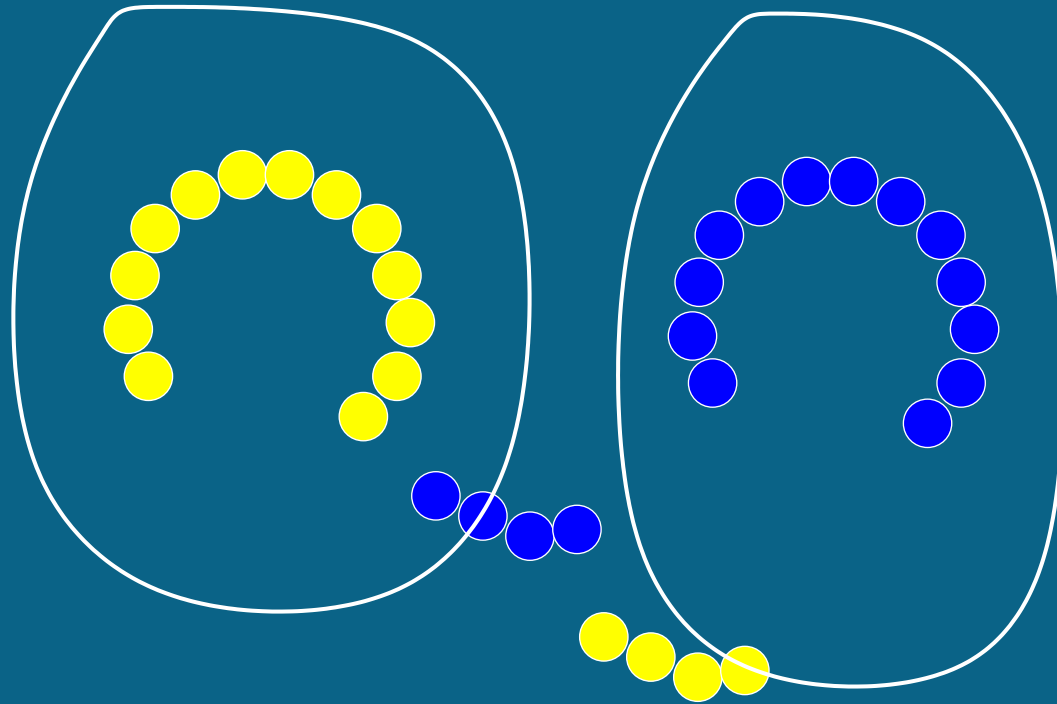
G

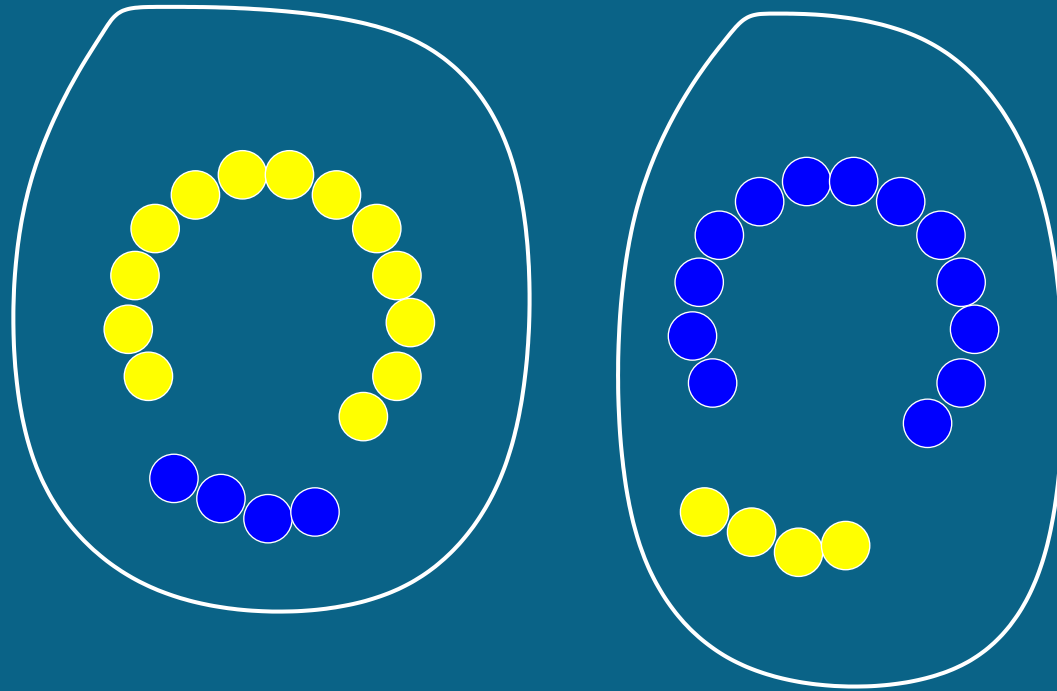
T

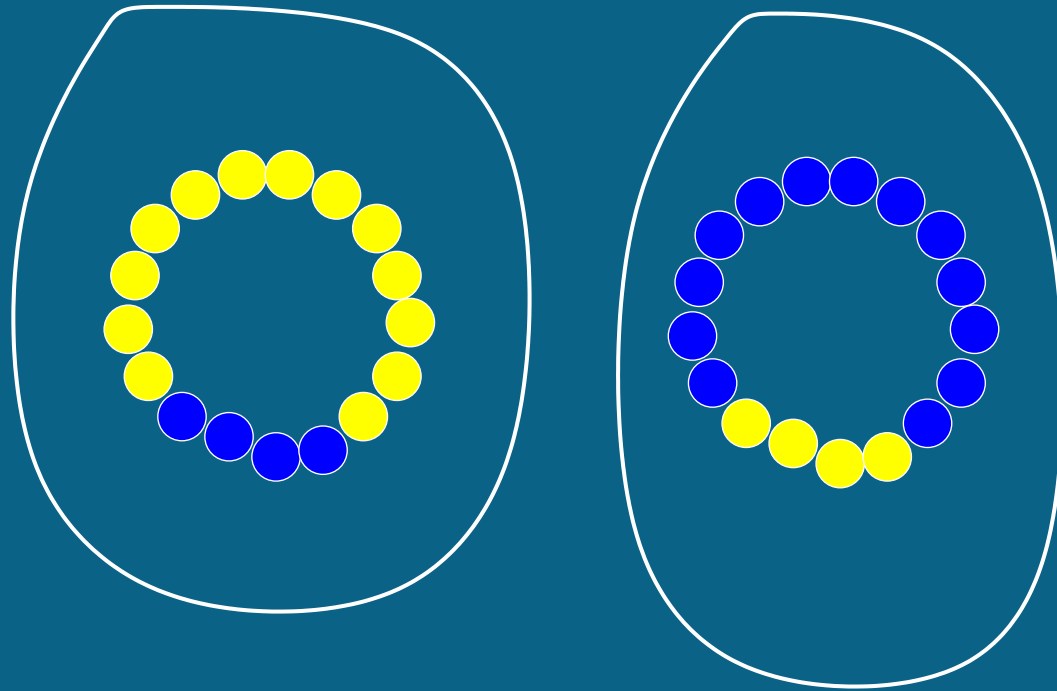


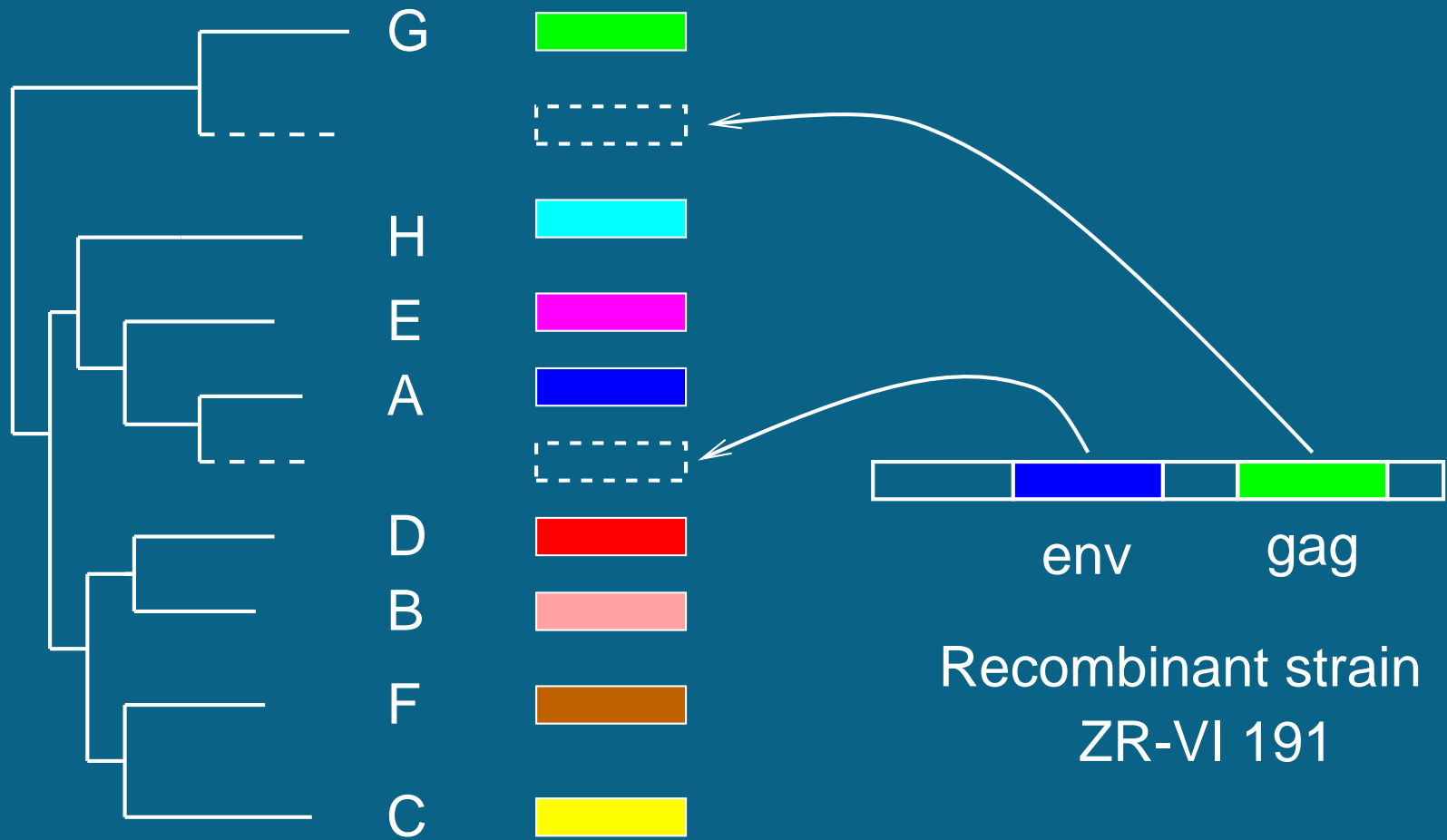


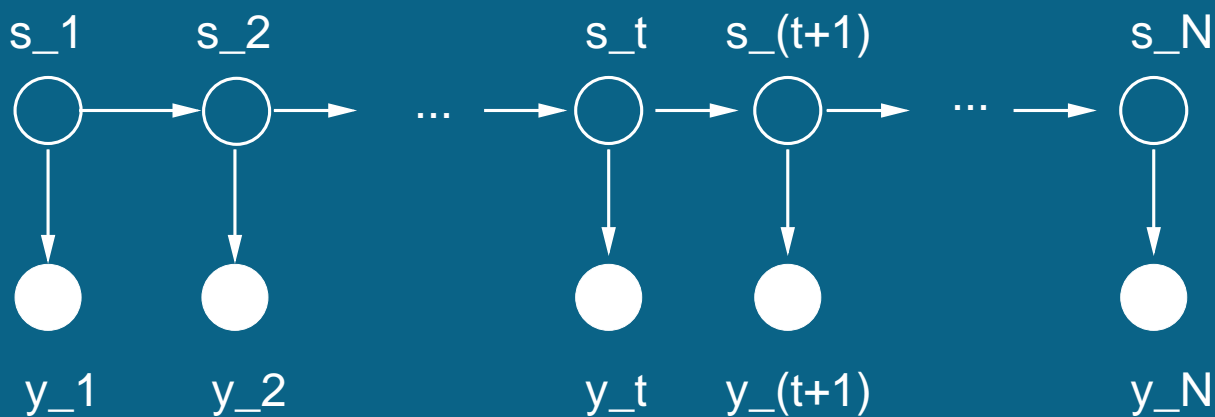
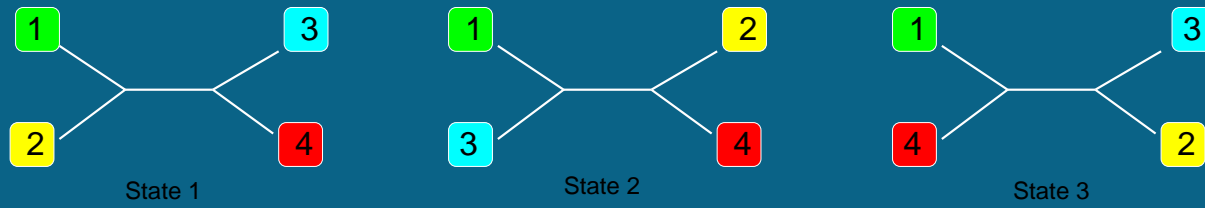






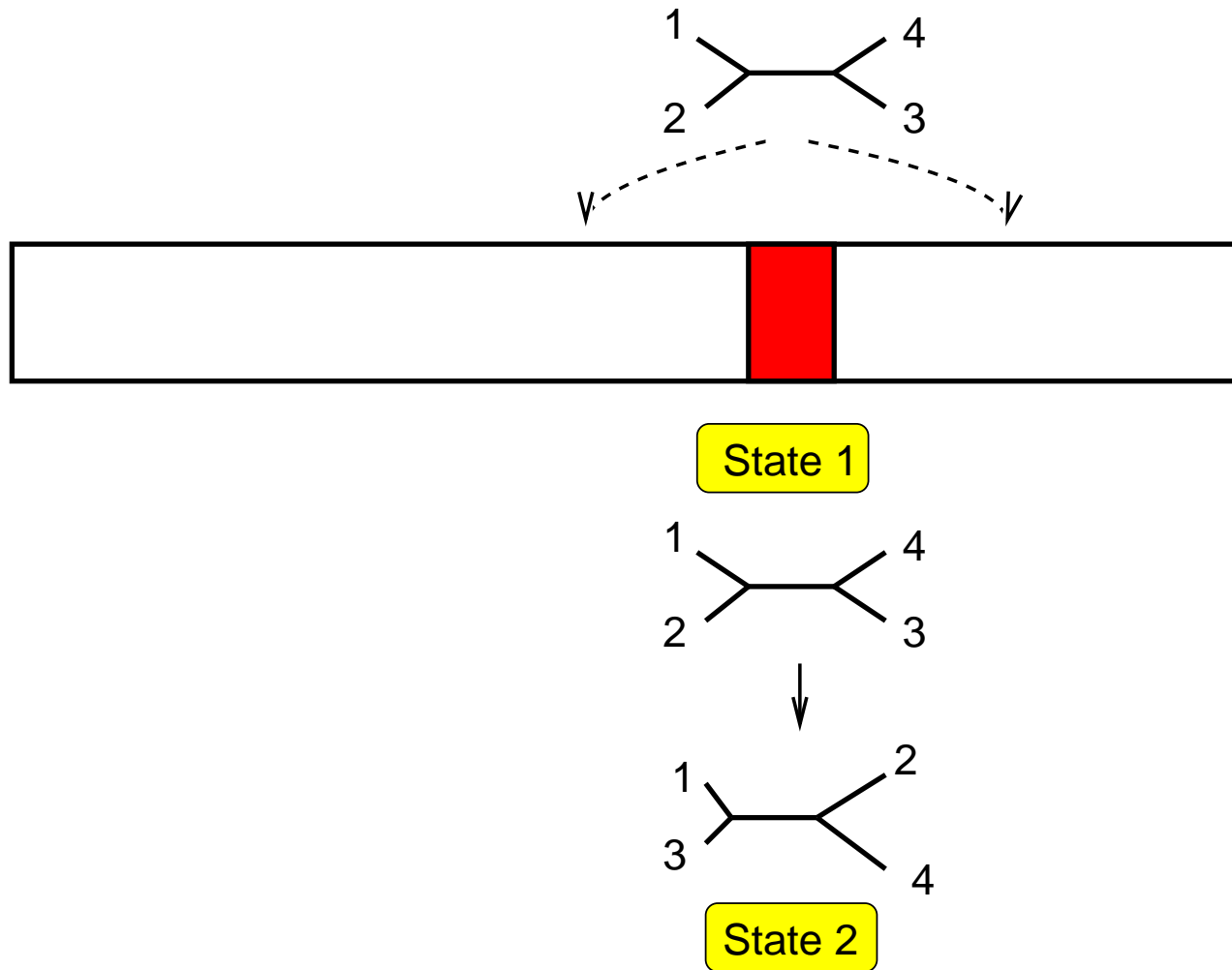




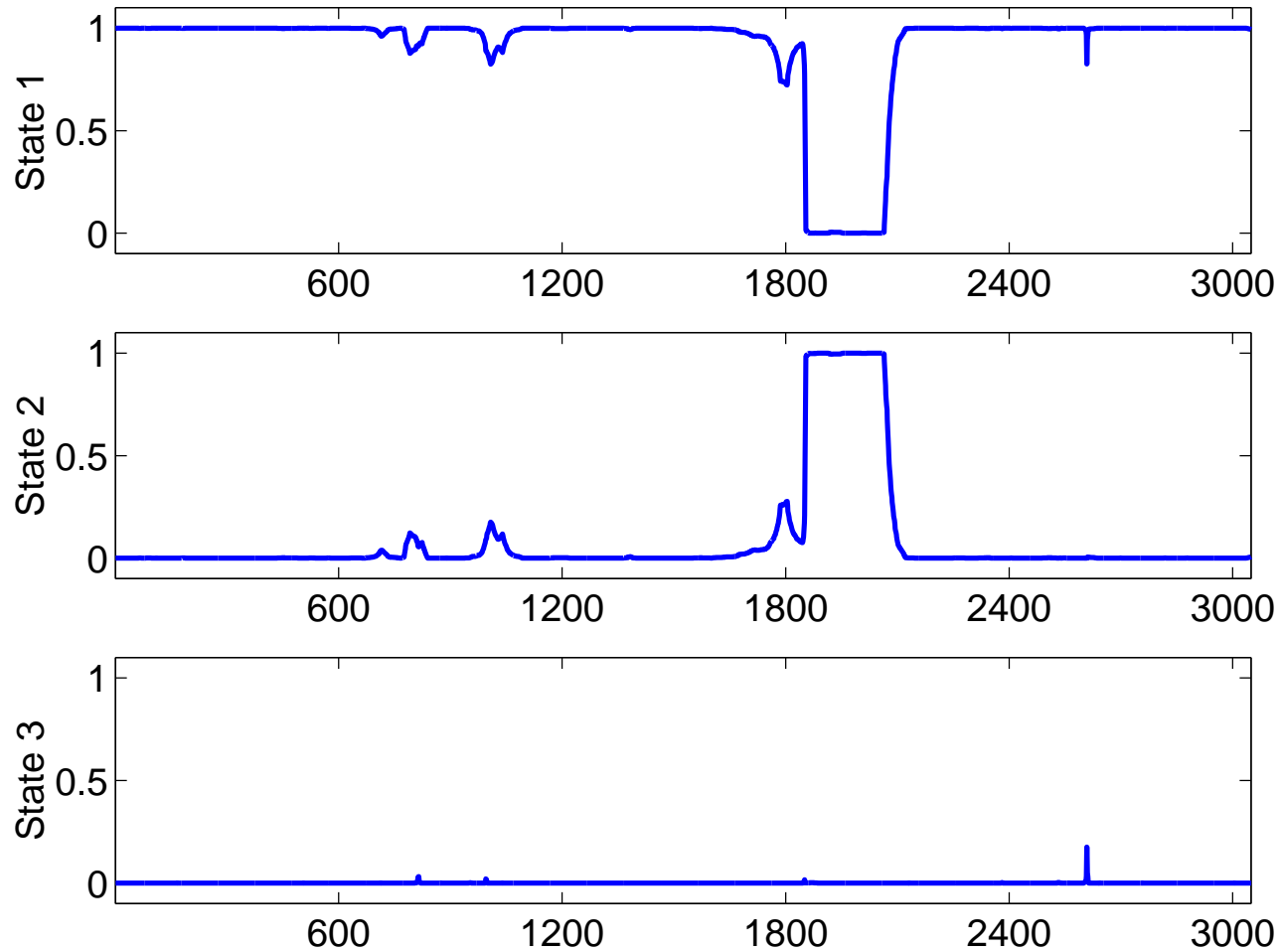


AGCATCGTTCTATTTTACCGGCTCCCG
 TGTGTCGCTCAAGATTGCCATCGCGCG
 TGTGTCGTGGTCTAGATTGCCATCGCGCG
 TGTATCGCTCTAGTTTGCCAGCTCCCG

4 strains of Hepatitis B virus: DNA alignment, 3049 nucleotides



4 strains of Hepatitis B virus: DNA alignment, 3049 nucleotides



Probabilistic Graphical Models

Fraud

Age

Sex

Gas

Jewelry

Fraud

Age

Sex

Gas

Jewelry

Fraud

Age

Sex

Gas

Jewelry

Fraud

Age

Sex

Gas

Jewelry

Fraud

Age

Sex

Gas

Jewelry

Fraud

Age

Sex

Gas

Jewelry

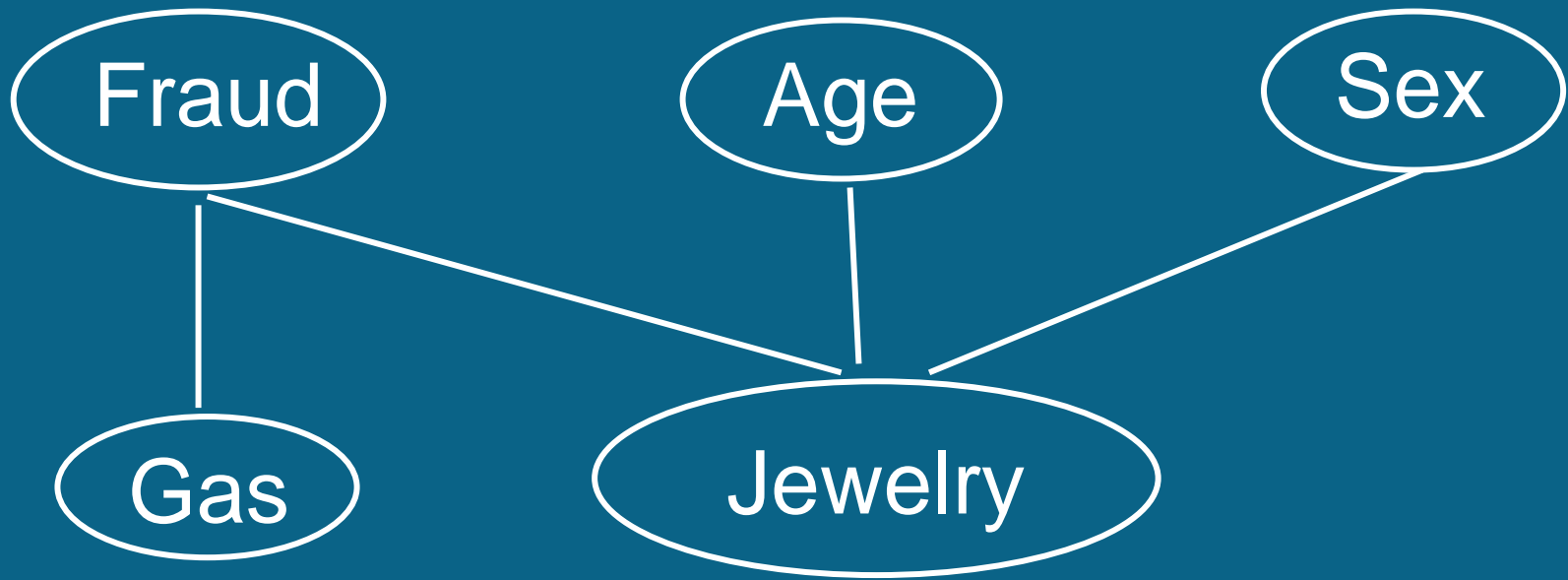
Fraud

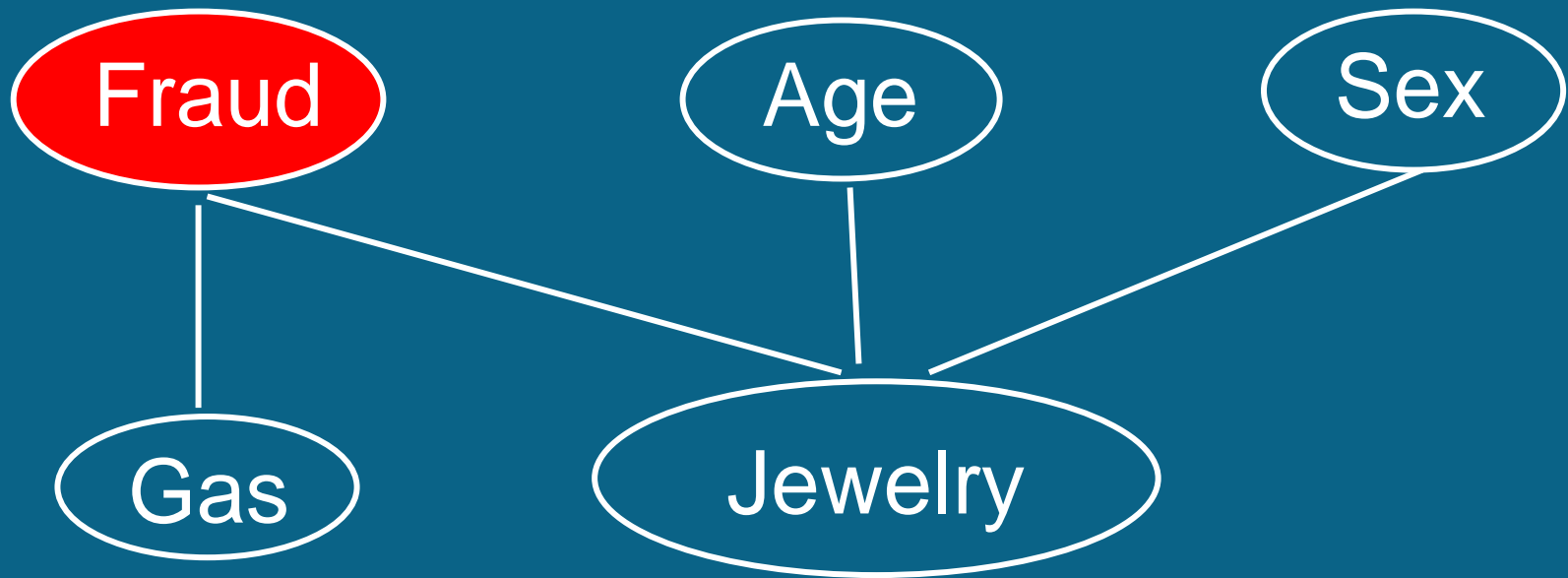
Age

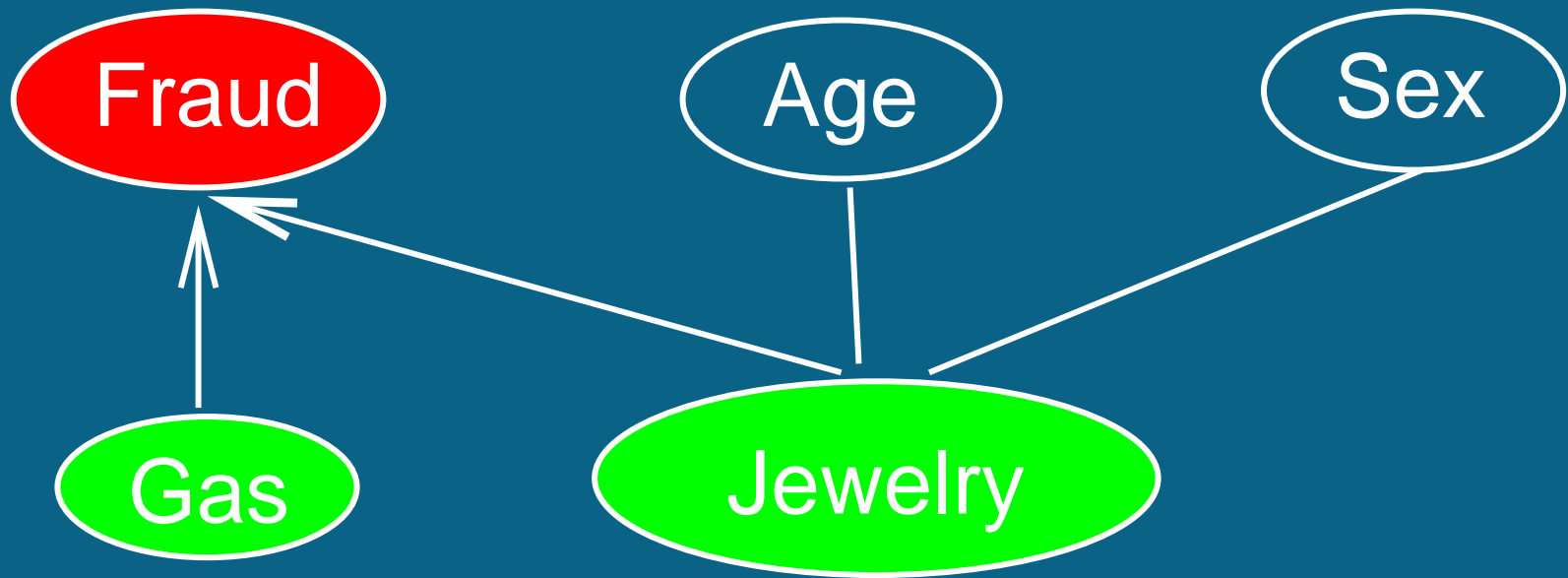
Sex

Gas

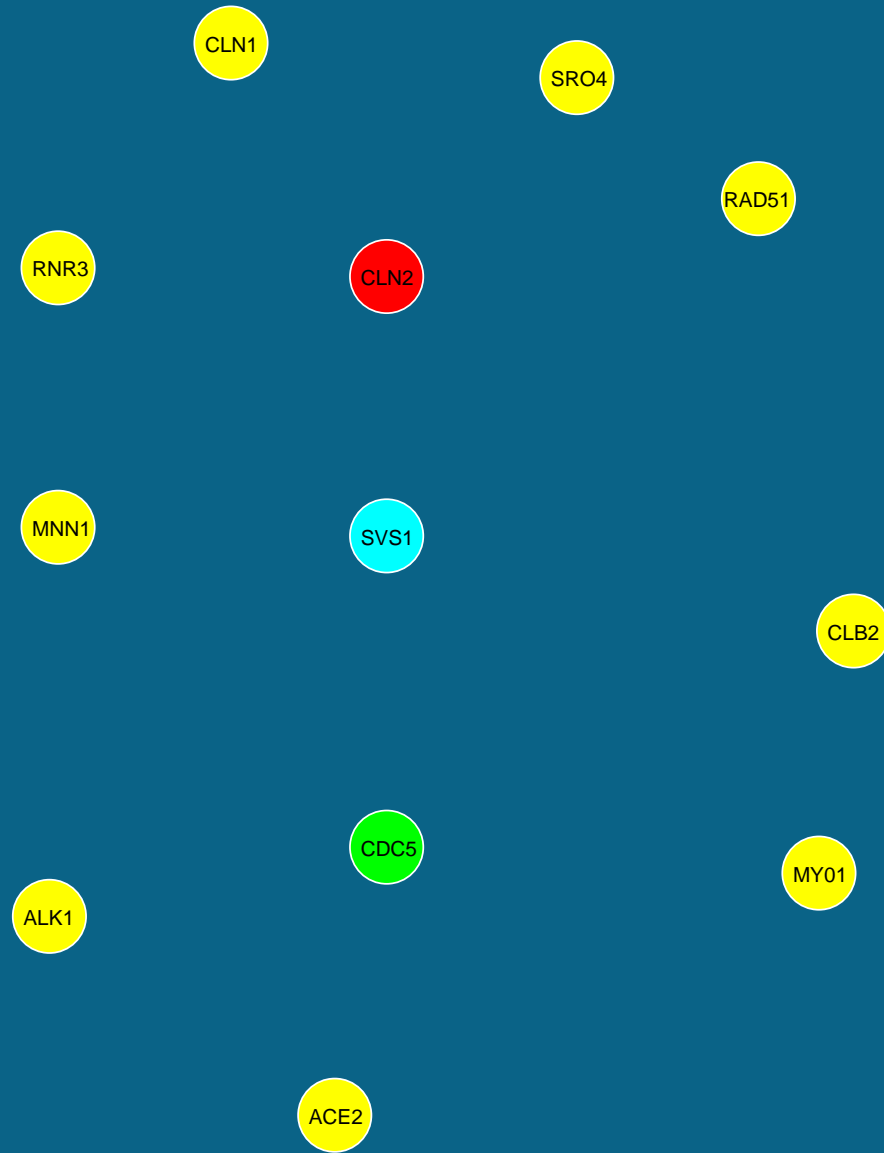
Jewelry

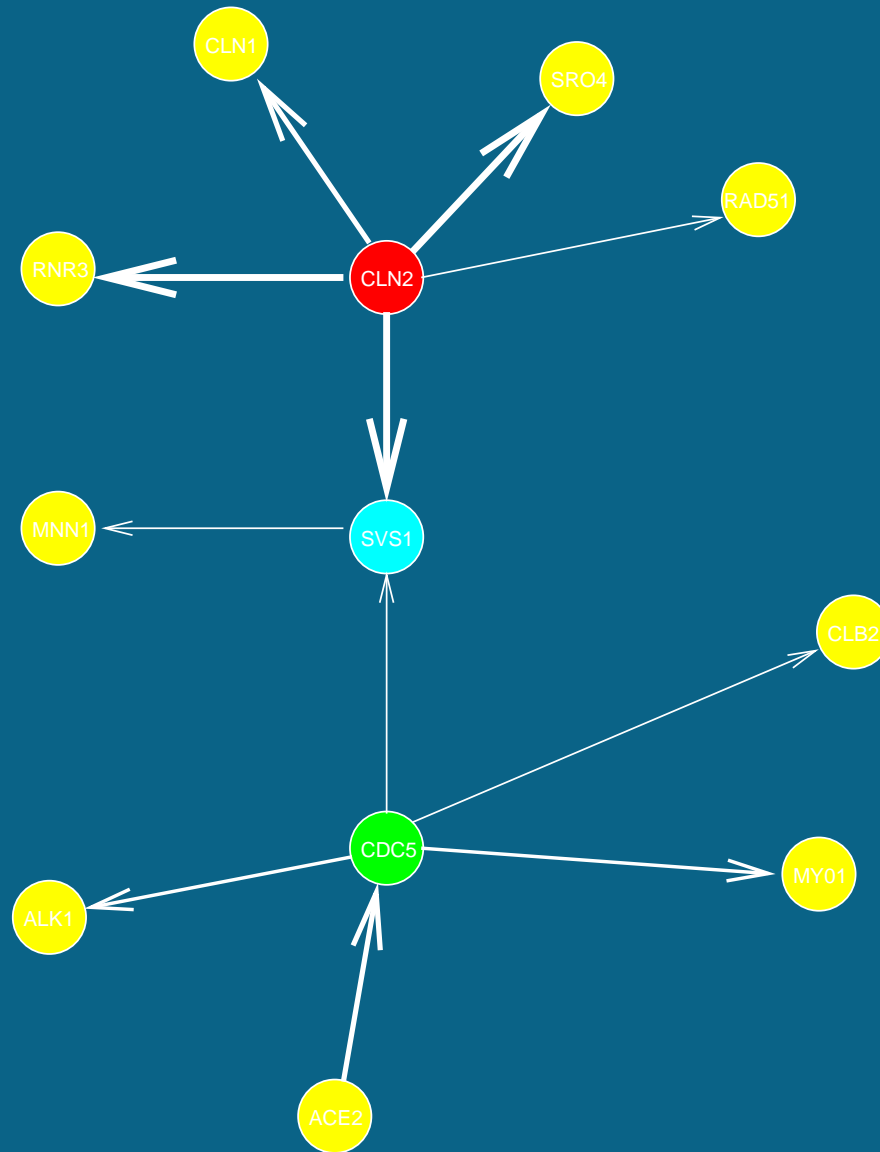






Application: Genetic Networks





Summary

Summary

- Machine learning techniques:

Summary

- Machine learning techniques:
 - Neural networks
 - Hidden markov models
 - Probabilistic graphical models

Summary

- Machine learning techniques:
 - Neural networks
 - Hidden markov models
 - Probabilistic graphical models
- Application when no 'physical' models exist.

Summary

- Machine learning techniques:
 - Neural networks
 - Hidden markov models
 - Probabilistic graphical models
- Application when no 'physical' models exist.
- Powerful nonlinear methods

Summary

- Machine learning techniques:
 - Neural networks
 - Hidden markov models
 - Probabilistic graphical models
 - Application when no 'physical' models exist.
 - Powerful nonlinear methods
 - Learning needs heuristics
-