Prediction of Protein Sumoylation Sites using Deep Learning

Known information

Sumoylated protein

K

Labeling
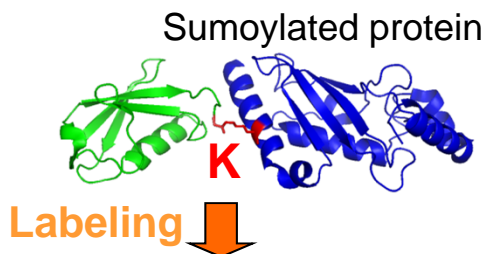
Positive data: sumoylation sites
Negative data: non-sumoylated lysines

Sequence space

MWCH**K**PADGGGE..SSFYFD**K**GGWGGG..
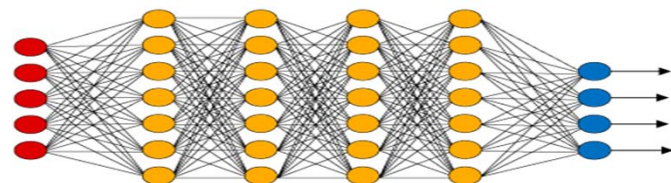**1**   **-1**
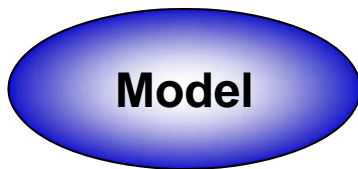
Input encoding

Biological Features

Feature space

6.0 7.0 8.3 6.0 **7.0** 7.0 7.0 .. 3.6 **7.0** 7.0 ..

Training

Classifier construction

Model

evaluation