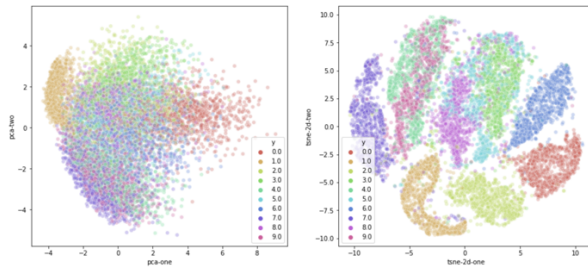
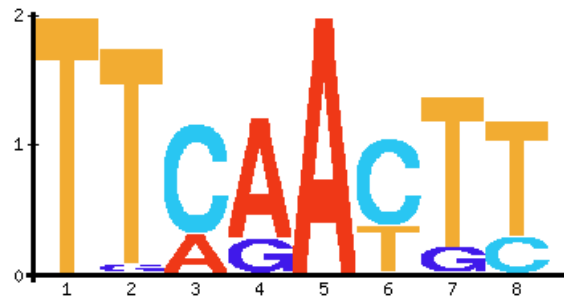
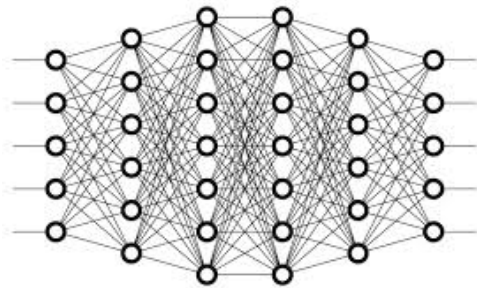


CS286: Topics in Sequence-based Machine Learning for Bioinformatics, Spring 2021

San Jose State University, Department of Computer Science
Professor William B. Andreopoulos

Tuesday-Thursday 4:30PM - 5:45PM (online)



```
CFIVGADNVGSKMQIRMSLRGK-AVVLGKNTMMRKAIRGHLENN--SALE
CFIVGADNVGSKMOTIRLSLRGK-AVVLGKNTMMRKAIRGHLENN--PALE
CFIVGADNVGSKMOTIRLSLRGK-AIVLMGKNTMMRKAIRGHLENN--PALE
CFIVGADNVGSKMONIRTSIRGL-AVVLGKNTMMRKAIRGHLENN--PALE
MIVAEADFGVSSQLQIRKSIIRGI-GAVLMGKNTMIRKVIIRDLADSK--PELD
MIVAEADFGVSSQLQIRKSIIRGI-GAVLMGKNTMIRKVIIRDLADSK--PELD
ILIVHVDNVGSKMASVRKSLRGK-ATILMGKNTIRRTALKKNLQAV--PQIE
IIIANIEGFPADKLHEIRKKLRGK-ADIKVTKNLFNIALKNAG----YDK
IIIANIEGFPADKLHEIRKKLRGK-AEIKVTKNLFNIALKNAG----LDVS
ILIGNLEGFPADKLHEIRKKLRGK-AEIKVTKNLFNIALKNAG----IDIE
VLFADLTGTFVVRVYRKKLWKK-YMMVAKKRILRAMKAAGLE---LDDN
VFLEDLHGLSIRLHEVRYRLRRY-GVIKTIKPLEKIAETKVYGG---TPAE
```

- Unsupervised/supervised Machine Learning in bioinformatics
- Logistic regression, Naïve Bayes
- Language models, Word embeddings
- Vector space representations
- Clustering with PCA, t-SNE, UMAP
- Hidden Markov Models and Markov chains
- Efficient sequence searching
- Recurrent Neural Networks for sequence modelling
- LSTMs for sequence prediction



011101011100101010010101010101010101001010101010100101001101100110100101011