Lecture 19: Text Generation with TensorFlow

LING 1340/2340: Data Science for Linguists

Ashley Feiler

Machine Learning Recap

Regression

- Predicting numerical values
- Supervised

Classification

- Predicting categorical values
- Supervised: Naïve Bayes, SVM, K-Nearest Neighbors
- Unsupervised: Clustering/Topic Modeling

The goal?

Language ANALYSIS

But what else is there?

Text Generation!

Markov Chain Models

Chat-GPT

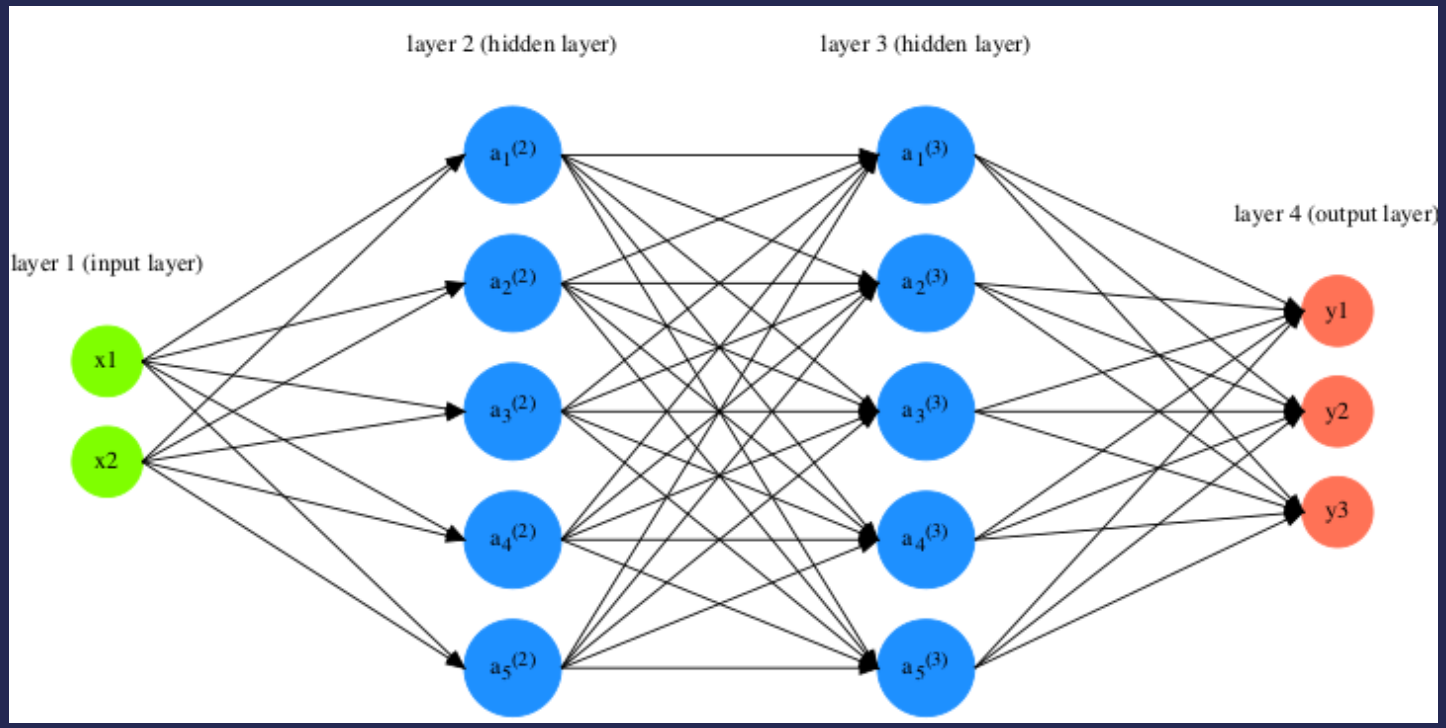
Recurrent Neural Networks

Google Gemini

Generative Pre-Training Transformers

Claude

RNNs: Recurrent Neural Networks



- Sequential data (ex. words)
- Has multiple hidden layers between input/output (deep learning)
- Remembers previous input by adjusting internal state
- Text-prediction task (supervised, but no annotations)

Why Deep Learning?

Data Scientist – NLP, LLM and GenAI

S&P Global
Virginia (+6 others)

- Hands-on experience with at least one current ML deep learning frameworks such as PyTorch and Tensorflow

AI/ML Researcher (LLM)

Alldus International Consulting Ltd

- Strong proficiency in programming languages such as Python, and experience with machine learning libraries/frameworks such as TensorFlow, PyTorch, or Hugging Face Transformers

Natural Language Processing Engineer at Petuum in Pittsburgh, PA

Petuum
Pittsburgh, PA

- Demonstrated hands-on experience with Python, Hugging Face, TensorFlow, Keras, PyTorch, Spark or similar statistical tools

Demo Time!

Shakespeare Demo Link