LECTURE 1: INTRODUCTION

DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING

Anatolii Stehnii
Software Developer in DataRobot

Yuri Guts
Software Developer in DataRobot

Lviv 2018

Ukrainian Catholic University
APPLIED SCIENCES FACULTY

WHAT YOU ALREADY HAVE LEARNED

- What is Natural Language Processing
- NLP levels
- NLP applications
- What are main NLP problems

MYSTERY OF LANGUAGE

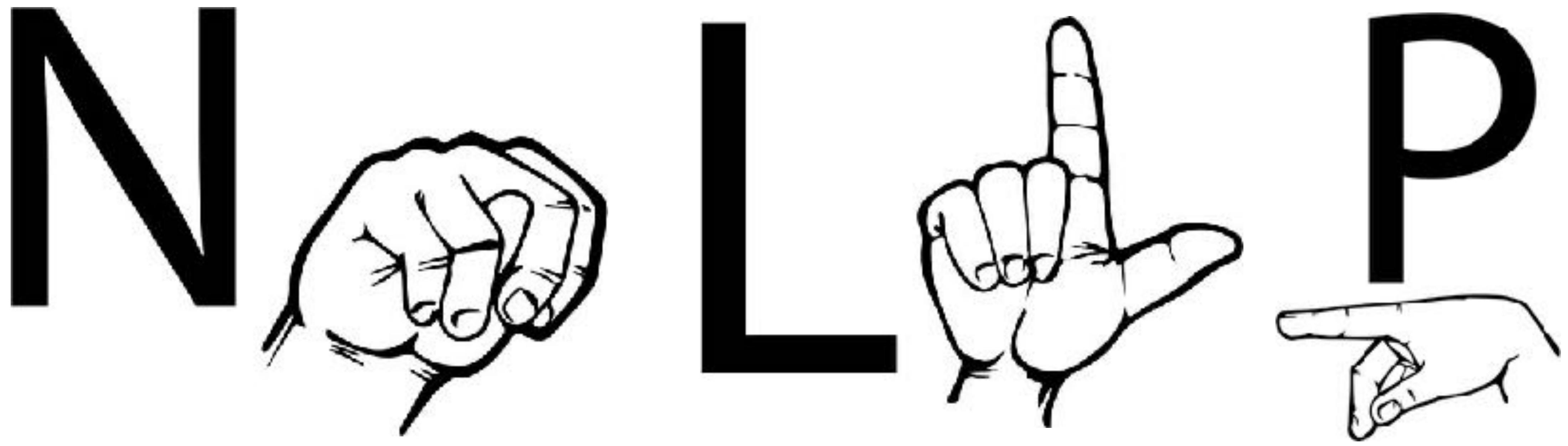
"… language is eccentric among animal communication systems … It has a core combination of features–semanticity, discrete infinity, and decoupling–that is found nowhere else in nature to our present knowledge … "

A Cognitive View of Language Evolution https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01434/full

Noam Chomsky: Cartesian Linguistics https://en.wikipedia.org/wiki/Cartesian_linguistics

LANGUAGE PROPERTIES

- **Discrete/symbolic/categorical** symbols: guitar -  gun - .
- **Naturally** evolved and contains multiple latent relations.
- Can have multiple **material encodings:** sound, writings, gesture.
- **Hierarchy** of symbols with steadily increasing degree of symbols complexity: letters, words, sentences, paragraphs, etc. **Infinite** creative possibilities.



MACHINE LEARNING AND NATURAL LANGUAGE

Discrete structure of natural language and it's large vocabulary leads to a **sparsity** of it's numeric representation.

Also, straightforward **language encoding** (BoW, TF-IDF) ignores **syntactic** and **semantic** information stored in a sentence structure. Feature engineering heavily relies on **human-designed** representations and requires a substantial expertise.

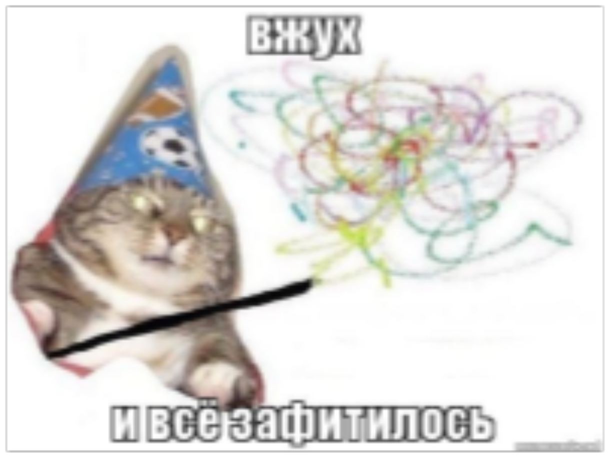
MACHINE LEARNING AND NATURAL LANGUAGE

Machine learning for NLP

Describing your data with features which can be used as algorithm inputs

Weight optimization

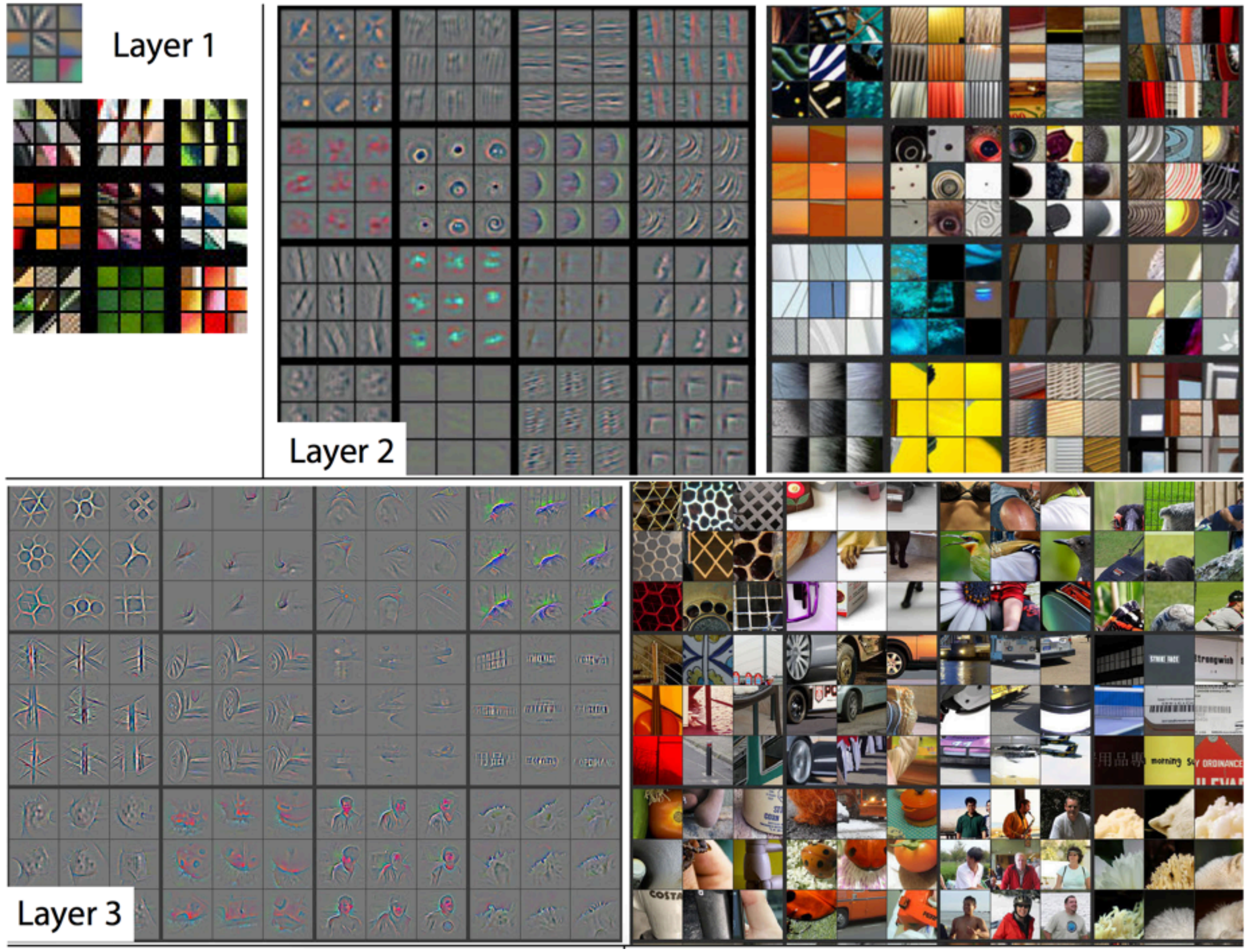
Domain specific, requires Ph.D. level expertise



DEEP LEARNING: REPRESENTATION LEARNING

Deep neural networks are composed from multiple **layers** (aka non-linear vector transformations), which are optimized by a supervision signal to provide domain specific hierarchical **representations (features)** of input data.

Deep learning can be used in NLP to avoid a process of complex manual feature engineering. **Syntactic** and **semantic** features can be incorporated by a NN from training data.



Zeiler & Fergus, 2013

NOT INDEPENDENT INPUTS

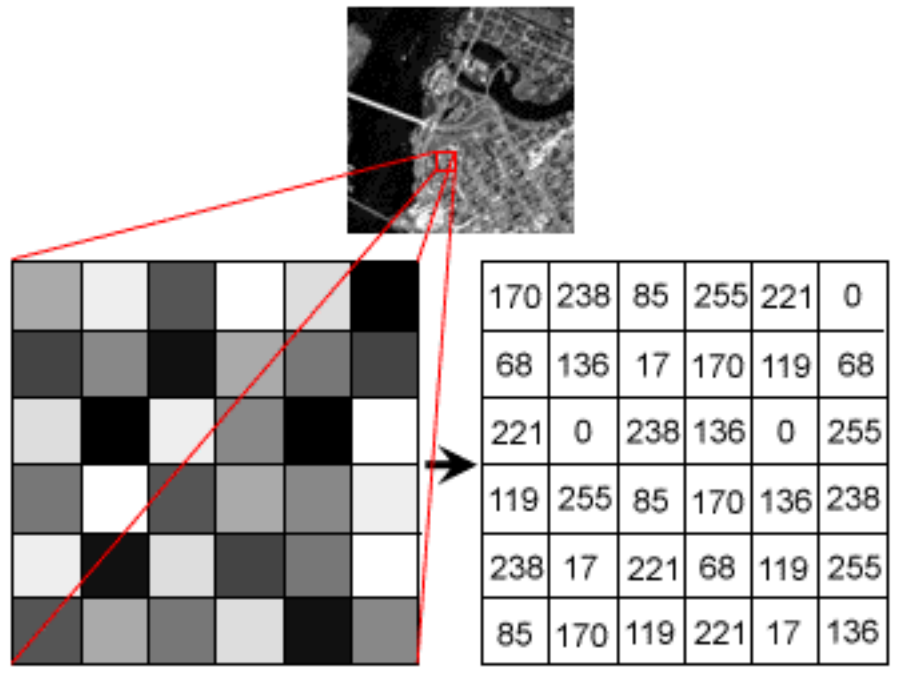
Classical machine learning algorithms, like logistic regression uses an statistical assumption, that their inputs are **independent random variables**.

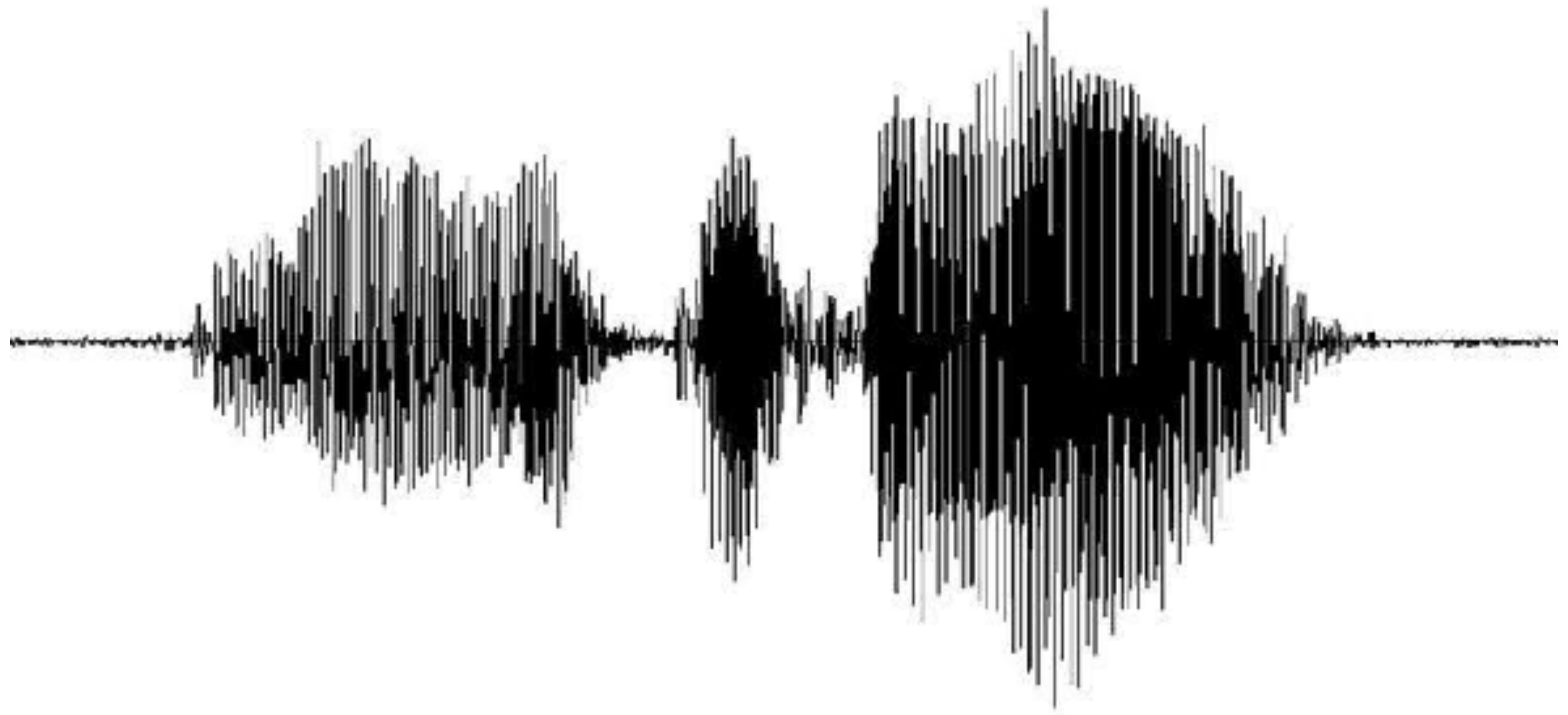
Low-level features of natural language (words, characters) are not independent. **Probability** of each word depends on it's context.

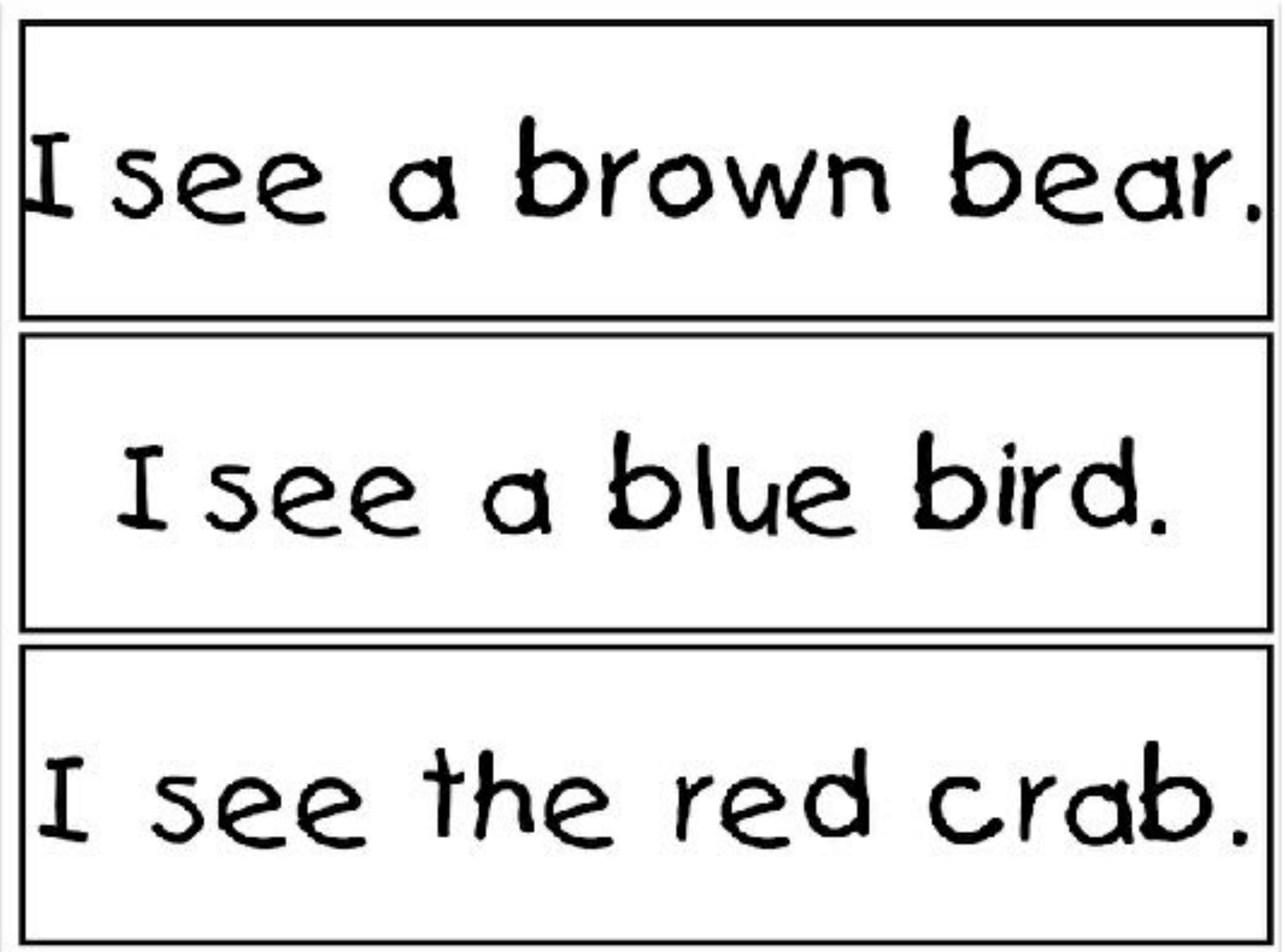
$$P(word, prev.words) = P(word|prev.words) \cdot P(prev.words)$$
$$P(word) \neq P(word|prev.words)$$

Q: What other examples of not independent data do you know?

NOT INDEPENDENT INPUTS

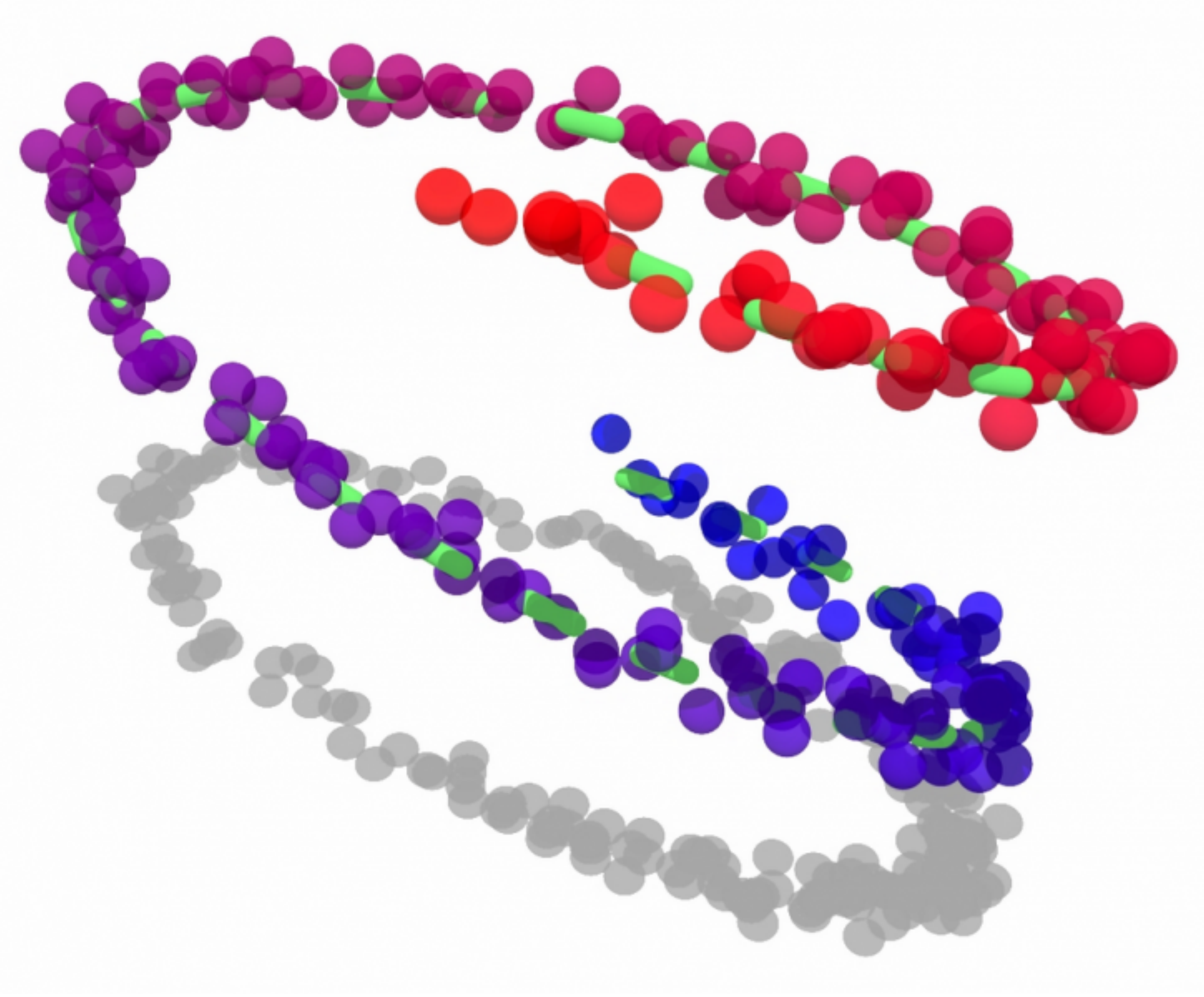






MANIFOLD HYPOTHESIS

Feature space of **cognitive information** (sound, image, text) are tremendously large (AlexNet input 227x227=51529 dimensions). However, only a tiny fraction of this space is used. Meaningful images are embedded in low-dimensional **manifolds.**



Zachary Pincus, 2010

DEEP LEARNING: MANIFOLD HYPOTHESIS

Deep learning techniques (like convolution or recurrent connection) assumes, that input data are **not distributed independently** over their feature space. During training, neural network **learns to unwrap** this manifolds and project data to a low-dimensional **representations.**

COURSE INFO

Teachers: Anatolii Stehnii, Yurii Guts.

Evaluation: practical assignments for each day.

Prerequisites: linear algebra, applied statistics, python (numpy).

Environment:

Python >= 3.5, pytorch >= 0.4.0, nltk, spacy, jupyter

COURSE STRUCTURE

Day 1

- Unsupervised deep learning for NLP
- Word embeddings

Day 2

- Convolution networks
- Recurrent networks
- Recursive networks

Day 3

- Sequence-to-sequence networks
- Attention and its friends
- Reinforcement learning for NLP