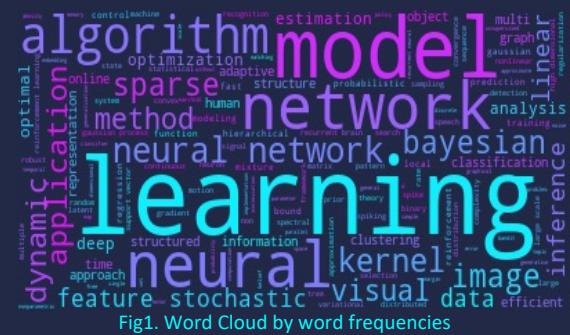


This project finds 5 hottest topics of the NIPS paper from 1987-2015. Amazingly, these topics can be identified as Computer Vision, Matrix Computation, Reinforcement Learning, Bayesian Methods and Time Series. Visualization outcomes of manifold methods strongly support the justification of topic extraction.

## 2. Data



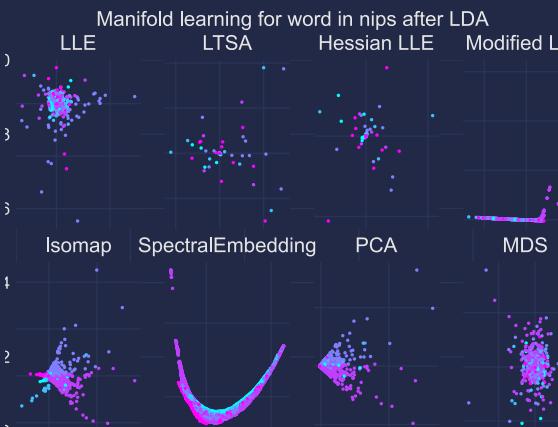
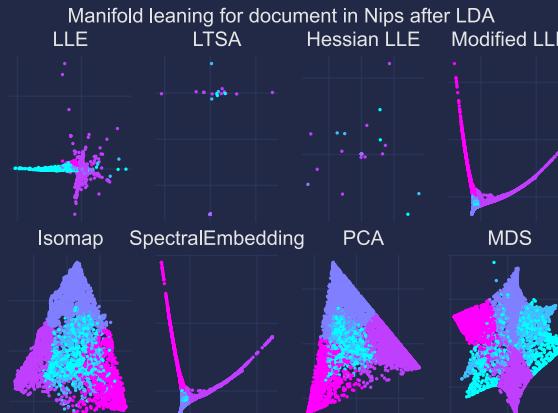
The dataset is a co-occurrence matrix for papers and words of size 11463 \* 5804  
Each row records a paper from NIPS 1987-2015  
Each column represents a word

### 3. Aims & Methodology

- Data Cleaning and Exploration. Showing Word Cloud, paper trends by year, topic trend by year.
  - Latent Dirichlet Allocation(LDA). Extracting 5 hottest topic from the data. Build the relationships of topics to each paper and word.
  - Manifold Learning. Visualizing papers and words with different topics.



We visualize paper with the topic it belongs and word with topic it contributes to respectively via 7 manifold methods and classical PCA and MDS.



More explanations and outcomes. Welcome to our GitHub! 

## 4. Data Exploration

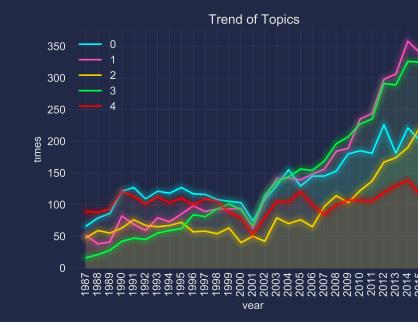


Fig2. Topic Trend by Year

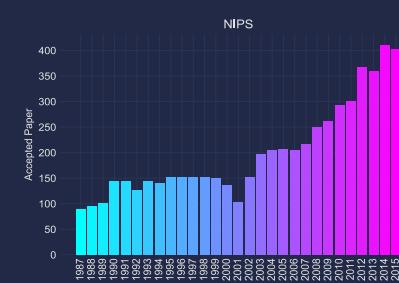


Fig3. Trend of the number of papers

## 7. Analysis & Conclusions

- given by LDA which preliminarily justifies our outcomes.
  - 2. Readers can tell specific hot topics like Computer Vision, Matrix Computation, Reinforcement Learning, Bayesian Methods and Time Series from the topic found by LDA .
  - 3. The beginning of this century is a turning point of the paper amount and topic amount, the number of them increase dramatically afterwards.
  - 4. Visualization shows clear patterns of paper-topic, but not so good of word-topic, which further validates our results.

## 8. References

**CSIC2011ADiveIntoNIPSWords Gu Hanlin, Huang Yifei, Sun Jiaze  
Y.Yao, A Mathematical Introduction to Data Science**

9. Contribution

ZHA Mengyue: Coding & Making the Poster  
HUANG Chutian: Theory Support & Writing the Paper