

Introduction to Deep Learning



Arijit Mondal

**Dept. of Computer Science & Engineering
Indian Institute of Technology Patna**

`arijit@iitp.ac.in`

List of Data Resources for Project

Some ideas: Computer vision

- Lung cancer detection - <https://www.kaggle.com/c/data-science-bowl-2017>
- Street view with house number - <http://ufldl.stanford.edu/housenumbers/>
- Image segmentation - <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>
- Image classification - <http://www.cs.toronto.edu/%7Ekriz/cifar.html>
- Visual question answering - <http://www.visualqa.org/>
- Image segmentation - <http://homepages.inf.ed.ac.uk/amos/afreightdata.html>
- Objects and scenes - <http://web.mit.edu/torralba/www/database.html>
- Open image data set - <https://github.com/openimages/dataset>
- Imagery feature detection - <https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection>

Some ideas: Natural language processing

- Daily news for stock market prediction - <https://www.kaggle.com/aaron7sun/stocknews>
- Question answering corpus - <https://github.com/deepmind/rc-data>, <http://cs.nyu.edu/%7Ekcho/DMQA/>
- Book review - <http://archive.ics.uci.edu/ml/datasets/Amazon+book+reviews>
- News popularity - <http://archive.ics.uci.edu/ml/datasets/Online+News+Popularity>
- Buzz in social media - <http://archive.ics.uci.edu/ml/datasets/Buzz+in+social+media+>
- Chat bot

Some ideas: Reinforcement learning

- Control of double inverted pendulum
- Adaptive traffic signal control

Some ideas: Regression

- Solar energy prediction - <https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest>
- Power consumption -
<http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>
- Optimal power flow -
<http://www.pjm.com/markets-and-operations/ops-analysis/historical-load-data.aspx>
- Modeling of flight delay -
https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/previous_years/

Some ideas: Others

- University ranking - <https://www.kaggle.com/mylesoneill/world-university-rankings>
- Car evaluation - <http://archive.ics.uci.edu/ml/datasets/Car+Evaluation>
- Common objects in context - <http://mscoco.org/home/>
- Youtube data set - <https://research.google.com/youtube8m/index.html>
- Gas sensor for home activity -
<http://archive.ics.uci.edu/ml/datasets/Gas+sensors+for+home+activity+monitoring>
- Online retail dataset - <http://archive.ics.uci.edu/ml/datasets/Online+Retail>
- Activity recognition from smart phone/watch -
<http://archive.ics.uci.edu/ml/datasets/Heterogeneity+Activity+Recognition>
- KDD cup 98 - <http://archive.ics.uci.edu/ml/datasets/KDD+Cup+1998+Data>
- KDD 2017 - Tollgate traffic flow prediction
- Music classification - http://marsyasweb.appspot.com/download/data_sets/
- Activity recognition - <https://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition>

NLP: liv.ai

- **Speech to text api, Audio search, etc.**
- **Supports 6 languages - English, Punjabi, Hindi, Telugu, Bengali, Kannada**
- **Chat client gappi**
- **Conversation system, chat bot, etc.**
- **Also, you can look api.ai**

Data resources

- <http://www.tdil-dc.in/>
- <http://www.nactem.ac.uk/resources.php>
- <http://www.openslr.org/12/>
- <http://www.ldcil.org/resourcesSpeechCorp.aspx>
- <http://www.statmt.org/europarl/>
- dbpedia
- <https://www.cs.cmu.edu/~cil/v-images.html>
- SIGKDD cup
- <https://www.kaggle.com/c/data-science-bowl-2018>
- <https://www.microsoft.com/en-us/research/event/interspeech-2018-special-session-low-resource-speech-recognition-challenge-indian-languages/>
- https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research
- <https://registry.opendata.aws/>

Project evaluation

- Need to form group with no more than 3 students (deadline - 22nd January)
- Abstract submission by 31st January
 - Title, brief description, dataset, etc.
- Final evaluation will be held in April before your end semester
 - At least one unique novel contribution
 - Innovative ideas
 - Within top 5 rank in Kaggle competition
 - Creation of dataset
 - Need to host code in github