



(<http://nlp.stanford.edu/>)

# CS224d: Deep Learning for Natural Language Processing

([index.html](#))

## Course Project

### Overview

The Course Project is worth a significant portion of your grade. It offers you the chance to flex your newly acquired skills toward an application of your choosing.

To inspire ideas, you might look at recent deep learning publications from top-tier NLP conferences and labs, as well as other resources below.

- NIPS (<http://nips.cc/>): Neural Information Processing Systems
- ICML (<http://icml.cc/>): International Conference on Machine Learning
- EMNLP (<http://emnlp2014.org>): Empirical Methods in Natural Language Processing
- ACL (<https://www.aclweb.org/>): Association for Computational Linguistics
- Past CS229 Projects (<http://cs229.stanford.edu/projects2013.html>): Example projects from Stanford machine learning class
- Stanford NLP Group (<http://nlp.stanford.edu/publications.shtml>): New and relevant papers from local faculty
- Kaggle challenges (<http://www.kaggle.com/>): An online machine learning competition website
- Stanford's CoreNLP (<https://github.com/stanfordnlp/CoreNLP>): A Java suite of Core NLP tools
- NLP+Twitter ([https://github.com/aritter/twitter\\_nlp](https://github.com/aritter/twitter_nlp)): Twitter NLP tools
- Chinese NLP tools (<https://github.com/NLPchina/nlp-lang>): NLP for Chinese text

For applications, this type of projects would involve careful data preparation, an appropriate loss function, details of training and cross-validation and good test set evaluations and model comparisons. Don't be afraid to think outside of the box.

For models, we have included below some pretty cool papers.

- **Natural Language Processing (Almost) From Scratch** (<http://www.jmlr.org/papers/volume12/collobert11a/collobert11a.pdf>)
- **LSTM: A Search Space Odyssey** (<http://arxiv.org/abs/1503.04069>)
- **Named Entity Recognition** (<http://www.aclweb.org/anthology/W09-1119>)
- **Part of Speech Tagging** (<http://nlp.stanford.edu/pubs/CICLing2011-manning-tagging.pdf>)
- **Coreference Resolution** (<http://www.eecs.berkeley.edu/~gdurrett/papers/durrett-klein-emnlp2013.pdf>)
- **CBOW and Skip-Gram** (<http://arxiv.org/pdf/1301.3781.pdf>)
- **Toward AI Complete Question Answering: A Set of Prerequisite Toy Tasks** (<http://arxiv.org/abs/1502.05698>)
- **Memory Networks** (<http://arxiv.org/pdf/1410.3916v8.pdf>)

If you would like to use AWS for your project, please read our tutorial (supplementary/aws-tutorial-2.pdf)

## Important Dates

Course project proposal: due April 21.

Course project milestone: due May 15.

The poster session will be held 2-5pm at Gates (AT&T patio) on June 1.

Final course project: due **June 3** (11:59pm).

## Grading Policy

```
Final Project: 40%
milestone: 5%
write-up: 10%
  • clarity, structure, language, references: 3%
  • background literature survey, good understanding of the problem: 3%
  • good insights and discussions of methodology, analysis, results, etc.: 4%
technical: 12%
  • correctness: 4%
  • depth: 4%
  • innovation: 4%
evaluation and results: 10%
  • sound evaluation metric: 3%
  • thoroughness in analysis and experimentation: 3%
  • results and performance: 4%
poster: 3% (+2% bonus for best few posters)
```

## Project Proposal

The project proposal should be a few short paragraphs (200-400 words overall). If you work on your own project, your proposal should contain the **headings**:

- **Problem Description:** What is the problem that you will be investigating? Why is it interesting?
- **Data:** What data will you use? If you are collecting new datasets, how do you plan to collect them?
- **Methodology/Algorithm:** What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations?
- **Related Work:** What reading will you examine to provide context and background?
- **Evaluation Plan:** How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

**Submission:** Please upload one proposal per team on Gradescope. Unless you have written a petition proposal to us (and we have accepted it) you are only allowed to have two people per team.

## Project Milestone

Your project milestone report should be between 2 - 3 pages using the provided template (project\_template/template.zip). The following is a suggested structure for your report:

- Title, Author(s)
- Introduction: this section introduces your problem, and the overall plan for approaching your problem
- Problem statement: Describe your problem precisely specifying the dataset to be used, expected results and evaluation
- Technical Approach and Models: Describe the methods you intend to apply to solve the given problem
- Intermediate/Preliminary Experiments & Results: State and evaluate your results upto the milestone

**Submission:** Please upload a on PDF file per team on Gradescope.

## Final Submission

Your final write-up should be between **6 - 8** pages using the provided template (project\_template/template.zip). After the class, we will post all the final reports online so that you can read about each others' work. If you do not want your writeup to be posted online, then please let us know at least a week in advance of the final writeup submission deadline.

Submit your final submission as instructed below:

1. A PDF file of your final report submitted through Gradescope.
2. (OPTIONAL) zip file with Supplementary Materials (e.g. code) through our Box folder.

Note that, each individual in a team is required to make submission (i.e. the same PDF and zip file) for grading purpose.

## Submit file to CS224D Staff

Please submit your supplementary files. Name the file with your SUID (e.g. jsmith.zip). Include a readme file.

---

File Upload: Choose File No file chosen

Description (255 characters maximum):

Your e-mail (required):

Upload

This upload widget is externally owned. Uploading files here will not upload them to your Box account.

powered by **box**

**Report.** The following is a suggested structure for the report:

- Title, Author(s)
- Abstract: It should not be more than 300 words
- Introduction: this section introduces your problem, and the overall plan for approaching your problem
- Background/Related Work: This section discusses relevant literature for your project
- Approach: This section details the framework of your project. Be specific, which means you might want to include equations, figures, plots, etc
- Experiment: This section begins with what kind of experiments you're doing, what kind of dataset(s) you're using, and what is the way you measure or evaluate your results. It then shows in details the results of your experiments. By details, we mean both quantitative evaluations (show numbers, figures, tables, etc) as well as qualitative results (show images, example results, etc).
- Conclusion: What have you learned? Suggest future ideas.
- References: This is absolutely necessary.

**Supplementary Material** is not counted toward your 6-8 page limit.

Examples of things to put in your supplementary material:

- Source code (if your project proposed an algorithm, or code that is relevant and important for your project.).
- Cool videos, interactive visualizations, demos, etc.

Examples of things to not put in your supplementary material:

- All of a submodules (Theano, Caffe, CoreNLP) source code.

- Various ordinary data preprocessing scripts.
- Any code that is larger than 1MB.
- Model checkpoints.
- A computer virus.

## Poster Session

We will hold a poster session in which you will present the results of your projects in form of a poster. The poster session will happen on June 3rd, 2:00-5:00pm, at AT&T patio (the lawn behind Gates building). Poster boards and easels will be provided.

## Example Project Reports

Your project reports should structure like a NLP conference paper (NIPS, ICML, EMNLP, ACL, etc.). You can find publications from Stanford NLP Group from here (<http://nlp.stanford.edu/publications.shtml>). In addition, you may also take a look at some previous projects from other Stanford CS classes, such as CS221 (<http://web.stanford.edu/class/cs221/sample-projects/>), CS229 (<http://cs229.stanford.edu/projects2013.html>), CS224W (<http://web.stanford.edu/class/cs224w/projects.html>) and CS231n (<http://web.stanford.edu/class/cs224w/projects.html>)

## Collaboration Policy

You can work in teams of up to **2** people.

## Honor Code

You may consult any papers, books, online references, or publicly available implementations for ideas and code that you may want to incorporate into your strategy or algorithm, so long as you clearly cite your sources in your code and your writeup. However, under no circumstances may you look at another group's code or incorporate their code into your project.

If you are doing a similar project for another class, you must make this clear and write down the exact portion of the project that is being counted for CS224d.