

**Wrap Up**

**Mausam**

# Grading

- 30% final exam
- 20% midterm
- 10% A1
- 12% A2
- 7% A3.1
- 8% A3.2
- 13% A4

# **Key Points of the Course**

# Key Points: NLP

- Challenges
  - Ambiguity, Ambiguity, Ambiguity, Sparsity
- Words: Morphology
- Sentences: Syntax
  - POS tagging, NP chunking, Parsing
- Sentences/Documents: Semantics
  - Words/Bigrams encode meanings, but are also sparse
  - Distributional Semantics, Shallow semantics
  - Patterns: bootstrapping
- Documents: Coreference, Discourse
- Applications
  - Information Extraction, Machine Translation, Summarization, Dialog

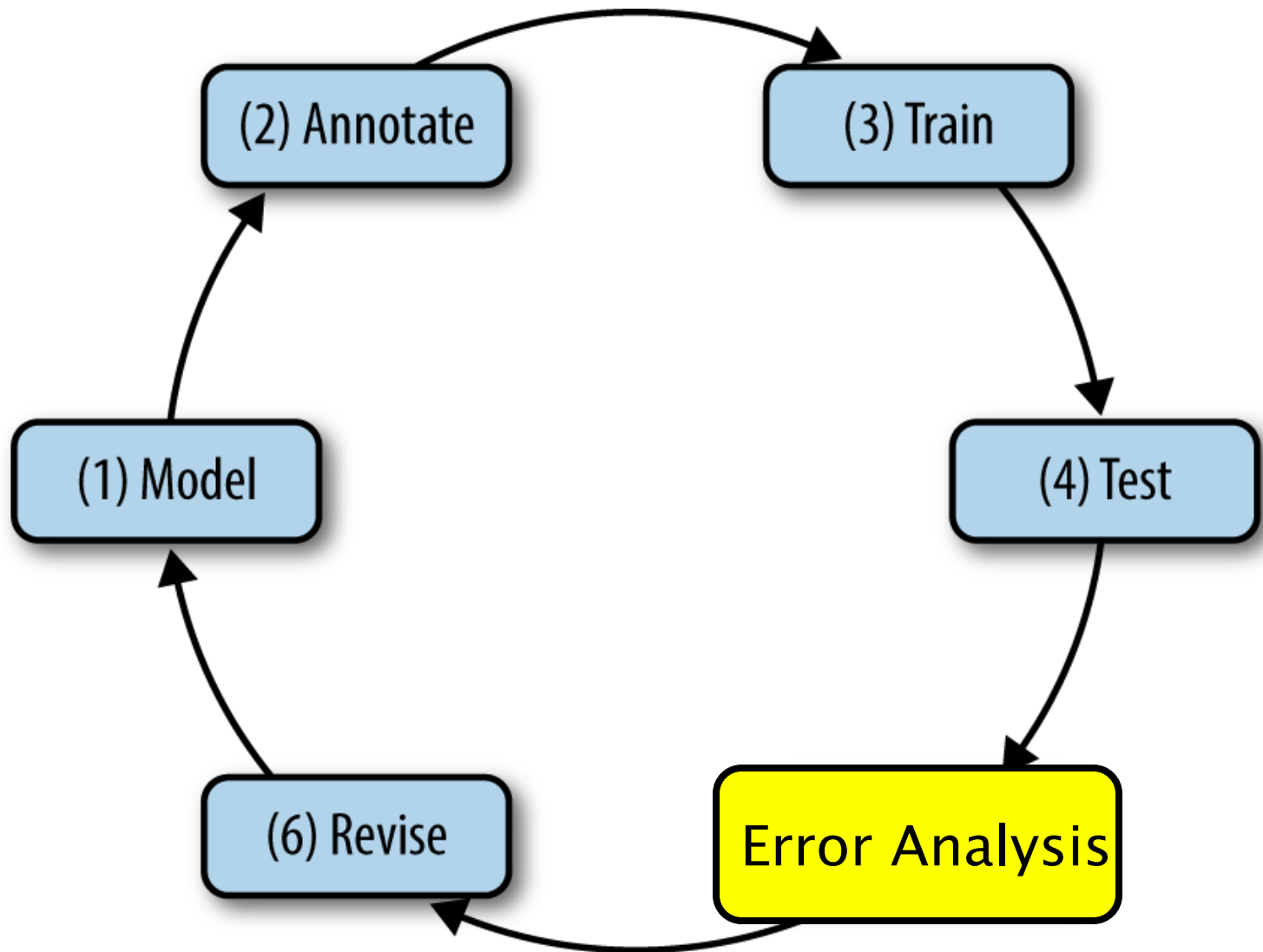
# Key Points: Modeling

- Models
  - Logical vs. Probabilistic vs. Neural
  - Representation: Bag of Word-level, Linear, Tree
  - Generative vs. Discriminative
  - Smoothing / Regularization
  - Pipeline vs. Joint inference
- Training Data
  - A lot: Supervised
  - None: Unsupervised/Self-supervised
  - A little: Semi-supervised
  - Distantly supervised
  - Active
- Features, Features, Features: local or global
- Learning Representations

# Key Points: Data Insights

- Data annotation: be a linguist!
- Crowdsourced Data Curation
- Data harvesting
  - Example: bootstrapping
  - Example: Summarization
- Indirectly related data
  - Example: distant supervision

# Key Points: ML Cycle



# Neural Models

- Shallow NNs
  - Bag(words)
- Convolutional NNs
  - Handle bag (fixed length n-grams)
- Recurrent NNs
  - Handle small variable length histories
- LSTMs/GRUs
  - Handle larger variable length histories
- Bi-LSTMs
  - Handle larger variable length histories and futures
- Recursive NNs
  - Handle variable length partially ordered histories



# Neural Models (contd)

- Neural language models
- Conditioned language models
  - Encoder-Decoder Models
- Attention models
  - attach non-uniform importance to histories based on evidence (question)
  - maybe even more important than recurrence?
- Transformer: all-word self attention + position embeddings
- Pre-trained language models
  - Fine-tuning
  - Light-weight finetuning: sidetuning, adapter
  - Prompt engineering

# **COL873: Advanced NLP**

# Details

- This will primarily be a paper reading and student presentation-driven course. The goal will be to learn about the state the art in NLP, learn how to read and critique research papers, try out research ideas, and write survey articles.
- Contents:
  - Pre-trained Language Models
  - Knowledge-Based NLP
  - Question Answering
  - Dialog Systems
  - Ethics in NLP
  - Machine Learning Ideas in NLP
- The midterm will require you to write a survey paper, and an open-book exam based on topics covered in the course. We will do an assignment and a project.