Wrap Up

Mausam

Grading

- 30% final exam
- 20% midterm
- 14% A1
- 14% A2
- 3% A3 (discussion)
- 19% A3

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

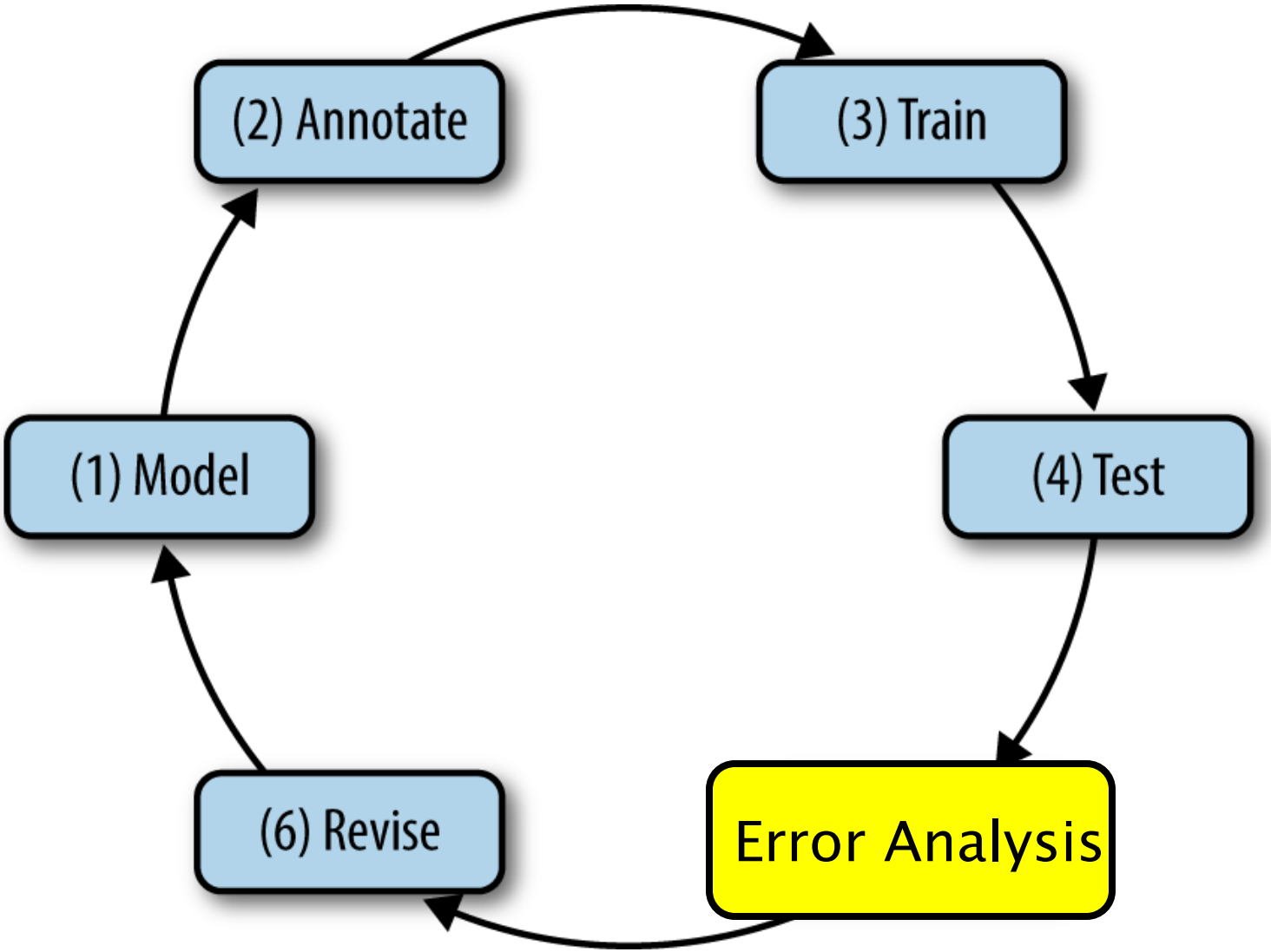
Key Points: Modeling

- 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

Low Data Setting

- Use GPTx zero-shot
- Use GPTx few-shot
- Use similar tasks; pretrain; then fine-tune few-shot
- Transfer learning; Multi-task learning
- Data augmentation
- Auxilliary loss
- Add human-features
- Add constraints
- Change architectures
- Research Questions: GPTx + Secondary model