January 10

Key points:
- In/dependent Probabilities
- Conditional Probabilities
- Bayes Rule
  - Elements in the equation: posterior, likelihood, prior

January 22

Key points:
- Bag-of-words representation of documents
- Maximum Likelihood Estimators
- Naive Bayes
  - Important property: assumes independence between different features (=words)

January 24
https://twitter.com/mjntendency/status/957667716684570625

Key points:
- Smoothing for Naive Bayes
- Grid search for best parameters over development set
- Bias/Variance tradeoff: more smoothing = more bias, for example
- Perceptron
  - Perceptron guaranteed to succeed if data is linearly separable
  - Weight update based on gradient of hinge loss

January 29

Key points:
- Learning by minimizing a loss function
- Batch versus online
- Logistic regression
  - Conditional probability $P(y \mid x)$
- Regularizing by penalizing squared norm of weights

Questions:
- Will NB and LR give the same answer if independence holds

January 31

Key points:
- Review of gradient based learning: examples for Perceptron, SVM, and Logistic Regression updates.
- Relationship between perceptron and logistic regression updates.
- Intro to deep learning

February 5

Key points:
- Ngram Language Models
- Neural Language Models (Recurrent Nets)

Questions:
- What problem is solved by the recurrent models?
  - Most exact answer: they learn a **probability**. In LMs, this is then used to predict the next word.

February 7

Key points:
- Evaluation of language models: perplexity
- Long Short-term Memory RNNs: memory "cell" in addition to hidden state
  - Less susceptible to vanishing gradients
- Part of speech tags

February 12

Key points:
- Decompose sequence scores into local scores that only look at adjacent tags
- **Viterbi algorithm** solves this form efficiently
- Tradeoff between efficiency and representational power
- Viterbi variables, trellis of scores, back-pointers

February 19

Key points:
- Structured Perceptron - compute argmax with Viterbi
- Conditional Random Field (CRF) - structured analog of logistic regression, maximize P(Y|X)
- Use Forward Algorithm to compute the denominator (marginal of X), forward variables
- Add margin regularizer to the objective - Structured SVM

Questions:

- Can we use anything other than Viterbi to compute the predictions in structured perceptron?
  - Approximation may be useful if you can't represent your features with exact inference

February 21

Key points:
- BiLSTM + CRFs: reverse recurrent models working together to represent states in the text
- Convolutional Neural Nets (ConvNets): take windows from input, then pool convolutional outputs
  - Fast on GPUs
- Named Entity Recognition (NER)
  - BIO annotation - beginning/inside/outside

Questions:
- Objectives which are probabilities are only necessary if their interpretation within the model is a probability.

February 26

Key points:
- Context Free Grammar (CFG) <Non-terminals, terminals, rules, start state> - defines a set of valid strings (language)
- Derivation/Parse Tree - from start state to the string - may have ambiguity
- Constituent
- Attachment Ambiguity

February 28

Key points:
- A derivation is a sequence of production rules
- Find derivations of a sentence with the CKY algorithm - start on the diagonal, then fill in the rest
  - (space) Complexity of CKY: $O(M^2*\#N)$, where $\#N$ = number of non-terminals
  - (time) complexity: $O(M^3*\#R)$, where $\#R$ = number of production rules
  - $\#R$ is probably larger than $\#N$
- Choose a derivation by taking the argmax, as usual, over scores of derivations
- Weighted context free grammar (W-CFG): attaches a score to each production rule
  - Can learn weights with log-probabilities, or discriminatively

March 12th

Key points:
- Headword for constituents (lexicalization in context-free grammar)
- Dependency tree graph - spanning trees - labeled edges
- Structure for dependency parsing (word, modifier, relation)

March 14th

Key points:
- Transition-based parsing: three possible actions (Shift, arc-left, arc-right)
- Arcs can have relation types
- An oracle converts a ground truth dependency tree into a sequence of actions that produces that tree
- More than one possible oracle due to "spurious ambiguity" - more than one sequence of actions can produce the same tree (example: 1 <- 2 -> 3)

March 26th

Key points:
- Graph-based parsing
- Lexical features and the benefits of using word embeddings
- Projectivity (all nodes between endpoints of the edge are descendants of them)
- Dependency paths

March 28th

Key points:
- Propositional Logic
- First Order Logic
- Lambda expressions
- Syntactic Semantic Grammar

April 2nd

Key points:
- "More complicated" Lambda expressions
- Building a semantic lexicon
  - Determiners - lambda-exps that operate on predicates
- Type raising
- Learning for logical form derivations
- Annotating data for semantic model learning

April 4th

Key points:
- Shallow semantics - expressiveness
- Comparison with First Order Logic
- Semantic role labelling

April 9th

Key points:
- Coreference resolution
    - Referring expressions, entities/events
- Different types of noun phrases behave differently in the coreference problem
    - Pronouns / nominals / proper nouns
- SMASH
    - Matching - using hard constraints on textual properties
    - Heuristics - syntactic parallelism, prominence
    - Finding headwords - parser / attention model

April 11th

Key points
- Word embeddings: real-valued vectors that represent words
- The Distributional Hypothesis - "similar words have similar contexts"
- Word2Vec -> CBOW,
- LSA - SVD of the word/context matrix
- Uses for word embeddings - downstream (neural models)

Questions
- How does w2v turn contextual information into vectors?

April 16th

Key points
- Entity Linking - match text to an entity in the knowledge base
    - Popularity and context as features
    - Vector representation for features and entities
    - Ranking loss
- Relation Extraction
    - Macro vs Micro reading
    - Micro reading strategy: link entities, look at the dependency parse (path between the entities)
    - Build RNN along the dependency path, max pooling

April 18th

Key points
- Systran: Rule Based Machine Translation
- Word ordering for translation is really difficult.
- Seq2Seq models (Encoder-Decoder)
    - Encoder encodes the source sentence representation
    - Decoder is a neural language model conditioned on the source representation
- Attention: Attention scores are used as weights and a weighted representation of the encoders hidden states is used as context vector for the decoder.
- Vauquois Pyramid
- Evaluation metrics: BLEU

Questions