How to build a POS tagger?

- Key sources of information:
  - 1. The word itself
  - 2. Word-internal characters
  - 3. POS tags of surrounding words: syntactic context

- Approach: supervised learning (text => tags)
  - Today/Thursday: with the Hidden Markov Model
  - Next week: Conditional Random Field (arbitrary features)
Sequence Tagging

- I saw Fred Smith go to the store
  - Named Entity Recog with "BIO" tagging
  - Part-of-Speech tagging

\[
\overrightarrow{W} \Rightarrow \overrightarrow{Y}
\]

Input \rightarrow Output
**Markov Model**

\[ w \rightarrow s \rightarrow w \quad P(w) = \Pi_t P(w_t | w_{t-1}) \]

**Hidden Markov Model**

\[ y_1 \rightarrow y_2 \rightarrow y_3 \rightarrow \cdots \]

\[ w_1 \quad w_2 \quad \frac{w_3}{y_3} \]

\[ P(w, y) = \Pi_t P(y_t | y_{t-1}) \frac{P(w_t | y_t)}{\text{Transition Model}} \]

\[ P(y_t | y_{t-1}) \frac{P(w_t | y_t)}{\text{Emission Model}} \]

\[ P(y_t | y_{t-1}) \frac{P(w_t | y_t)}{\text{Observation Model}} \]

**Other Examples**

- Music
- Economics \( y = \text{Recession?} \) \( w = \text{Econ Stats} \)
- Radar Tracking \( y = \text{Ship loc.} \) \( w = \text{Obs. signal} \)
HMM uses \( P(\mathbf{w}, \mathbf{y}) = T T^t P(\mathbf{w} | \mathbf{y}_t) P(y_{t+1} | y_{t-1}) \)

Learning: Supervised Learning with \((\mathbf{w}^r, y^r)\)

Inference

\(1\) Likelihood (LM)

\[ P(\mathbf{w}) = \sum_{\mathbf{y}} P(\mathbf{w}, \mathbf{y}) \]

Effec.

Alg: Forward Alg.

\(2\) Tagging/Prediction: \( \arg\max_{\mathbf{y}} P(\mathbf{y} | \mathbf{w}) \)

Effec.

Alg: Variable Algo.