Today’s Lecture

- Discriminative Models
  - Logistic Regression
  - Maximum Entropy Markov Model
  - Conditional Random Field

- Named Entity Recognition
Recap – Models

- Generative – $P(x, y)$
- Discriminative – $P(y|x)$
Recap – Naive Bayes

\[
\arg\max_y P(y|x) = \arg\max_y P(y) P(x|y) \\
\approx \arg\max_y P(y) \prod_i P(x_i|y)
\]
Recap – Naive Bayes

\[
\arg\max_y P(y|x) = \arg\max_y P(y) P(x|y) \\
\approx \arg\max_y P(y) \prod_i P(x_i|y)
\]
Recap – Naive Bayes

\[
\text{argmax}_{y} P(y|x) = \text{argmax}_{y} P(y) P(x|y)
\]

\[
\approx \text{argmax}_{y} P(y) \prod_{i} P(x_i|y)
\]

Discriminative – approximate \( P(y|x) \)?
Logistic Regression

\[ P(y|x) \approx \frac{1}{Z} \exp \left( \sum_i \theta_{y,i}x_i \right) \]
Logistic Regression

\[ P(y|x) \approx \frac{1}{Z} \exp \left( \sum_i \theta_{y,i} x_i \right) \]

\[ = \frac{\exp \left( \sum_i \theta_{y,i} x_i \right)}{\sum_{y'} \exp \left( \sum_i \theta_{y',i} x_i \right)} \]
Logistic Regression

\[
P(y|x) \approx \frac{1}{Z} \exp \left( \theta_y + \sum_i \theta_{y,i}x_i \right)
\]

\[
= \frac{\exp \left( \theta_y + \sum_i \theta_{y,i}x_i \right)}{\sum_{y'} \exp \left( \theta_{y'} + \sum_i \theta_{y',i}x_i \right)}
\]
Logistic Regression

\[ P(y|x) \approx \frac{1}{Z} \exp \left( \theta_y + \sum_i \theta_{y,i} x_i \right) \]

\[ = \frac{\exp \left( \theta_y + \sum_i (\theta_{y,i} + k) x_i \right)}{\sum_{y'} \exp \left( \theta_{y'} + \sum_i (\theta_{y',i} + k) x_i \right)} \]
Logistic Regression

\[ P(y|x) \approx \frac{1}{Z} \exp \left( \theta_y + \sum_i \theta_{y,i} x_i \right) \]

\[ = \frac{e^{kx_i} \exp \left( \theta_y + \sum_i \theta_{y,i} x_i \right)}{e^{kx_i} \sum_{y'} \exp \left( \theta_{y'} + \sum_i \theta_{y',i} x_i \right)} \]
Logistic Regression

\[ P(y|x) \approx \frac{1}{Z} \exp \left( \theta_y + \sum_i \theta_{y,i}x_i \right) \]

\[ = \frac{\exp \left( \theta_y + \sum_i \theta_{y,i}x_i \right)}{\sum_{y'} \exp \left( \theta_{y'} + \sum_i \theta_{y',i}x_i \right)} \]
Naive Bayes

\[ y \]

\[ x_1 \]
\[ x_2 \]
\[ x_3 \]
\[ x_4 \]
\[ x_5 \]
Logistic Regression

\[ y \]
Logistic Regression

- Parameters: $\theta_y, \theta_{y,i}$

No closed form formula!
Logistic Regression

- Parameters: $\theta_y, \theta_{y,i}$
- Optimise for: $\sum_{(x,y) \in D} \log P(y|x)$
Logistic Regression

- Parameters: \( \theta_y, \theta_{y,i} \)
- Optimise for: \( \sum_{(x,y) \in D} \log P(y|x) \)
- No closed form formula!
Independence of Features

$x_1$  $x_2$  $x_3$  $x_4$  $x_5$
Independence of Features

- Hong Kong vs. HongKong
Independence of Features

- Hong Kong vs. HongKong
- Naive Bayes:
  - $P(x_i|y)$ same
  - $P(y|x)$ over-estimated
Independence of Features

- Hong Kong vs. HongKong

- Naive Bayes:
  - $P(x_i|y)$ same
  - $P(y|x)$ over-estimated

- Logistic Regression:
  - $P(y|x)$ same
  - $P(x_i|y)$ never used!
Logistic Regression

\[ P(y|x) \approx \frac{1}{Z} \exp \left( \sum_i \theta_{y,i} x_i \right) \]
Why Log-Linear?

- Consider all distributions $P(y|x)$
Why Log-Linear?

- Consider all distributions $P(y|x)$
- Under constraints:
  - $P(y|x_i)$ matches observed data
Why Log-Linear?

- Consider all distributions $P(y|x)$

- Under constraints:
  - $P(y|x_i)$ matches observed data

- Maximise conditional entropy $H(Y|X)$ on observed data
Why Log-Linear?

- Consider all distributions $P(y|x)$
- Under constraints:
  - $P(y|x_i)$ matches observed data
- Maximise conditional entropy $H(Y|X)$ on observed data

→ Logistic regression
Regularisation

- Equivalent of smoothing
- Optimise objective function:

\[ \mathcal{L} = \log P(y|x) - \lambda |\theta| \]
Recap: Hidden Markov Model

$t_1$ -> $t_2$ -> $t_3$ -> $t_4$

$w_1$ -> $w_2$ -> $w_3$ -> $w_4$
MaxEnt Markov Model

$\begin{align*}
&t_1 \\
&\downarrow \\
&\quad W_1 \\
&t_2 \\
&\downarrow \\
&\quad W_2 \\
&t_3 \\
&\downarrow \\
&\quad W_3 \\
&t_4 \\
&\downarrow \\
&\quad W_4
\end{align*}$
MaxEnt Markov Model

- MaxEnt: logistic regression
- Markov: limited context
MaxEnt Markov Model

- MaxEnt: logistic regression
- Markov: limited context
- Locally normalised: token by token
MaxEnt Markov Model

- MaxEnt: logistic regression
- Markov: limited context
- Locally normalised: token by token
- Dynamic programming for inference
Conditional Random Field

\[
\begin{align*}
t_1 & \rightarrow t_2 \\
t_2 & \rightarrow t_3 \\
t_3 & \rightarrow t_4 \\
W_1 & \rightarrow W_2 \\
W_2 & \rightarrow W_3 \\
W_3 & \rightarrow W_4
\end{align*}
\]
Conditional Random Field

- Conditional: discriminative
- Random field: undirected
Conditional Random Field

- Conditional: discriminative
- Random field: undirected
- Globally normalised: all at once
Conditional Random Field

- Conditional: discriminative
- Random field: undirected
- Globally normalised: all at once
- Dynamic programming or beam search for inference
Bill Gates says mosquitoes scare him more than sharks.
Bill Gates says mosquitoes scare him more than sharks.
Bill Gates says mosquitoes scare him more than sharks.

The reaction will produce 2,4- and 2,6-dinitrotoluene.
Bill Gates says mosquitoes scare him more than sharks.

The reaction will produce 2,4- and 2,6-dinitrotoluene.
Named Entity Recognition

- Sequence labelling task
- Usually into classes: PER, LOC, etc.
Bill Gates says mosquitoes scare him more than sharks.

B beginning
I inside
O outside
Bill Gates says mosquitoes scare him more than sharks

beginning

inside

outside
Bill Gates says mosquitoes scare him more than sharks.
Bill Gates says mosquitoes scare him more than sharks

B beginning
I inside
O outside
Bill Gates says mosquitoes scare him more than sharks.
Bill Gates says mosquitoes scare him more than sharks.

B I O O

scare him more than sharks

O O O O O

B beginning
I inside
O outside
Bill Gates says mosquitoes scare him more than sharks.

- Bill: B-PER
- Gates: I-PER
- says: O
- mosquitoes: O
- scare: O
- him: O
- more: O
- than: O
- sharks: O

B beginning
I inside
O outside
Defining the Task

The New York Stock Exchange fell today.
Defining the Task

The **New York Stock Exchange** fell today.
Defining the Task

The New York Stock Exchange fell today.
Defining the Task

The New York Stock Exchange fell today.
The New York and Chicago Stock Exchanges fell today.
Defining the Task

Queen Elizabeth
the Queen
the Queen of England
the queen of England
a queen of England
the queen of France
Features for Named Entity Recognition

- Gazeteers (lists of names)
- Capitalisation
- Digits
- Punctuation
- Specific words preceding/following (Prof., Inc.)