Introduction to data science & artificial intelligence (INF7100)

Arthur Charpentier

#434 Sequential Observations

été 2020

Markov Chains

Markov

20,000 caracters from *Eugène Onegin*, by Alexandre Pouchkine

Heedless of the proud world's enjoyment, I prize the attention of my friends, and only wish that my employment could have been turned to worthier ends ...

$$\begin{cases} P_{11} = \mathbb{P}[\text{letter}_{i+1} \in \text{voyel} | \text{letter}_i \in \text{voyel}] \\ P_{12} = \mathbb{P}[\text{letter}_{i+1} \in \text{conson} | \text{letter}_i \in \text{voyel}] \\ P_{21} = \mathbb{P}[\text{letter}_{i+1} \in \text{voyel} | \text{letter}_i \in \text{conson}] \\ P_{22} = \mathbb{P}[\text{letter}_{i+1} \in \text{conson} | \text{letter}_i \in \text{conson}] \end{cases}$$

In Russian, Andrej Markov obtained

$$P = \begin{pmatrix} 12,8\% & 87,2\% \\ 66,3\% & 33,7\% \end{pmatrix},$$

$\pi = (43,2\% \ 56,8\%)$ are respective frequencies of voyels and conson. In French $\pi = (45,6\% \ 54,4\%)$, in Italian $\pi = (47,4\% \ 52,6\%)$, and German $\pi = (38,5\% \ 61,5\%)$, see First Links in the Markov Chain.

Markov Chains

Consider sequential observations $X_1, X_2, \cdots$.
Markov property (of order 1) means

$$\mathbb{P}(X_{t+1} = x | X_t = x_t, X_{t-1} = x_{t-1}, \cdots) = \mathbb{P}(X_{t+1} = x | X_t = x_t)$$

$$\begin{array}{ccccccccccc}
\text{e} & & & \text{o} & \text{y} & & \boxed{\text{e}} & \rightarrow & \boxed{?} & 12.8\% \\
 & \text{n} & \text{j} & & & \text{m} & & \searrow & \boxed{?} & 87.2\%
\end{array}$$

$$\begin{array}{cccccccccccc}
\text{e} & & & \text{o} & \text{y} & & \text{e} & & \nearrow & \boxed{?} & 66.3\% \\
 & \text{n} & \text{j} & & & \text{m} & & \boxed{\text{n}} & \rightarrow & \boxed{?} & 33.7\%
\end{array}$$

If $x \in \mathcal{X}$, the transition matrix can be represented as a matrix

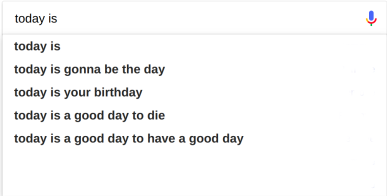
From letters to words

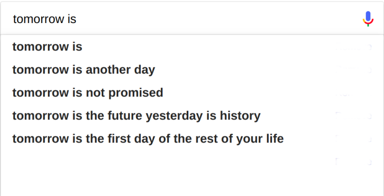
See Natural Language Processing) by Nivre

Chain Rule:

$$\mathbb{P}[A_1, A_2, \cdots, A_n] = \mathbb{P}[A_1] \cdot \mathbb{P}[A_2|A_1] \cdots \mathbb{P}[A_n|A_1, \cdots, A_{n-1}]$$

$$\mathbb{P}[A_1, A_2, \cdots, A_n] = \prod_{i=1}^{n} \mathbb{P}[A_i|A_1, A_2, \cdots, A_{i-1}] \simeq \prod_{i=1}^{n} \underbrace{\mathbb{P}[A_i|A_{i-1}]}_{\text{bigrams}}$$

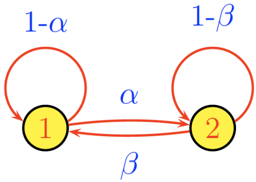




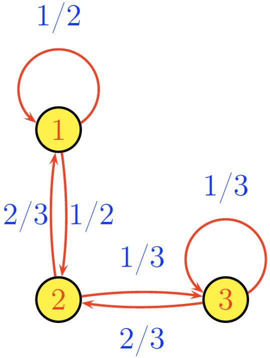
Markov Chains & Graphs

We've seen transition probabilities
(sum over all raws should equal to 1)
Example

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 2/3 & 0 & 1/3 \\ 0 & 2/3 & 1/3 \end{pmatrix}$$



We can visualize them on graphs or networks

See also Google's PageRank