Neural networks

Conditional random fields - performing classification

INFERENCE

Topics: computing $p(\mathbf{y}|\mathbf{X})$

- Then we have:

$$p(\mathbf{y}|\mathbf{X}) = \exp\left(\sum_{k=1}^{K} a_u(y_k) + \sum_{k=1}^{K-1} a_p(y_k, y_{k+1})\right) / Z(\mathbf{X})$$

where

$$Z(\mathbf{X}) = \sum_{y'_1} \sum_{y'_2} \cdots \sum_{y'_K} \exp\left(\sum_{k=1}^{K} a_u(y'_k) + \sum_{k=1}^{K-1} a_p(y'_k, y'_{k+1})\right)$$

hard to compute

INFERENCE

Topics: computing $p(y_k|\mathbf{X}),\ p(y_k,\ y_{k+1}|\mathbf{X})$

- The α / β tables can be used to compute marginals

$$p(y_k|\mathbf{X}) = \frac{\exp(a_u(y_k) + \log\alpha_{k-1}(y_k) + \log\beta_{k+1}(y_k))}{\sum_{y'_k}\exp(a_u(y'_k) + \log\alpha_{k-1}(y'_k) + \log\beta_{k+1}(y'_k))}$$

$$p(y_k, y_{k+1}|\mathbf{X}) = \frac{\exp\left(\begin{array}{c} a_u(y_k) + a_p(y_k, y_{k+1}) + a_u(y_{k+1}) \\ + \log\alpha_{k-1}(y_k) + \log\beta_{k+2}(y_{k+1}) \end{array}\right)}{\sum_{y'_k}\sum_{y'_{k+1}}\exp\left(\begin{array}{c} a_u(y'_k) + a_p(y'_k, y'_{k+1}) + a_u(y'_{k+1}) \\ + \log\alpha_{k-1}(y'_k) + \log\beta_{k+2}(y'_{k+1}) \end{array}\right)}$$

CLASSIFICATION

Topics: making a prediction (option 1)

- At each position k, pick label y_k with highest marginal probability $p(y_k|\mathbf{X})$
  - this choice is the one that minimizes the sum of the classification errors over the whole sequence, assuming the CRF is the true distribution

$$\min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right]$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X})
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X})
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_1} \cdots \sum_{y_K} 1_{y_k \neq y_k^*} \; p(\mathbf{y}|\mathbf{X})
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_1} \cdots \sum_{y_K} 1_{y_k \neq y_k^*} \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; \sum_{y_1} \cdots \sum_{y_{k-1}} \sum_{y_{k+1}} \sum_{y_K} p(\mathbf{y}|\mathbf{X})
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_1} \cdots \sum_{y_K} 1_{y_k \neq y_k^*} \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; \sum_{y_1} \cdots \sum_{y_{k-1}} \sum_{y_{k+1}} \sum_{y_K} p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; p(y_k|\mathbf{X})
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_1} \cdots \sum_{y_K} 1_{y_k \neq y_k^*} \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; \sum_{y_1} \cdots \sum_{y_{k-1}} \sum_{y_{k+1}} \sum_{y_K} p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; p(y_k|\mathbf{X}) \\
= \quad & \min_{y_1^*} \left(\sum_{y_1} 1_{y_1 \neq y_1^*} \; p(y_1|\mathbf{X})\right) + \cdots + \min_{y_K^*} \left(\sum_{y_K} 1_{y_K \neq y_K^*} \; p(y_K|\mathbf{X})\right)
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*} \right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*} \right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_1} \cdots \sum_{y_K} 1_{y_k \neq y_k^*} \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; \sum_{y_1} \cdots \sum_{y_{k-1}} \sum_{y_{k+1}} \sum_{y_K} p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; p(y_k|\mathbf{X}) \\
= \quad & \min_{y_1^*} \left(\sum_{y_1} 1_{y_1 \neq y_1^*} \; p(y_1|\mathbf{X}) \right) + \cdots + \min_{y_K^*} \left(\sum_{y_K} 1_{y_K \neq y_K^*} \; p(y_K|\mathbf{X}) \right) \\
= \quad & \min_{y_1^*} (1 - p(y_1^*|\mathbf{X})) + \cdots + \min_{y_K^*} (1 - p(y_K^*|\mathbf{X}))
\end{aligned}
$$

$$
\begin{aligned}
& \min_{\mathbf{y}^*} \mathrm{E}\left[\sum_k 1_{y_k \neq y_k^*}\right] \\
= \quad & \min_{\mathbf{y}^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_{y_1} \cdots \sum_{y_K} \left(\sum_k 1_{y_k \neq y_k^*}\right) \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_1} \cdots \sum_{y_K} 1_{y_k \neq y_k^*} \; p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; \sum_{y_1} \cdots \sum_{y_{k-1}} \sum_{y_{k+1}} \sum_{y_K} p(\mathbf{y}|\mathbf{X}) \\
= \quad & \min_{y_1^*} \ldots \min_{y_K^*} \sum_k \sum_{y_k} 1_{y_k \neq y_k^*} \; p(y_k|\mathbf{X}) \\
= \quad & \min_{y_1^*} \left(\sum_{y_1} 1_{y_1 \neq y_1^*} \; p(y_1|\mathbf{X})\right) + \cdots + \min_{y_K^*} \left(\sum_{y_K} 1_{y_K \neq y_K^*} \; p(y_K|\mathbf{X})\right) \\
= \quad & \min_{y_1^*} (1 - p(y_1^*|\mathbf{X})) + \cdots + \min_{y_K^*} (1 - p(y_K^*|\mathbf{X})) \\
= \quad & 1 - \max_{y_1^*} p(y_1^*|\mathbf{X}) + \cdots + 1 - \max_{y_K^*} p(y_K^*|\mathbf{X})
\end{aligned}
$$

CLASSIFICATION

Topics: making a prediction (option 2)

- Find most probable prediction:

$$\arg\max_{\mathbf{y}^*} p(\mathbf{y}|\mathbf{X})$$

- A Viterbi decoding algorithm can be used for that
  - forward pass: replace $\log \sum_{y'_k} \exp(\cdot)$ by $\max_{y'_k}(\cdot)$ and fill the table
  - backward pass: starting from the maximal value y^*_K at the last position, follow the backward trace of the max operations to decode the maximizing sequence

CLASSIFICATION

Topics: making a prediction (option 2)

nb. of classes

Complexity in $\mathrm{O}(KC^2)$

- Algorithm goes as follows:
  - initialize, for all values of y'_2 :
    - $\alpha^*_1(y'_2) \Longleftarrow \max_{y'_1} a_u(y'_1) + a_p(y'_1, y'_2)$
    - $\alpha^{\leftarrow}_1(y'_2) \Longleftarrow \arg\max_{y'_1} a_u(y'_1) + a_p(y'_1, y'_2)$
  - for $k = 2$ to K-1, for all values of y'_{k+1} :
    - $\alpha^*_k(y'_{k+1}) \Longleftarrow \max_{y'_k} a_u(y'_k) + a_p(y'_k, y'_{k+1}) + \alpha^*_{k-1}(y'_k)$
    - $\alpha^{\leftarrow}_k(y'_{k+1}) \Longleftarrow \arg\max_{y'_k} a_u(y'_k) + a_p(y'_k, y'_{k+1}) + \alpha^*_{k-1}(y'_k)$
  - $\max_{\mathbf{y}^*} p(\mathbf{y}|\mathbf{X}) \Longleftarrow \max_{y'_K} a_u(y'_K) + \alpha^*_{K-1}(y'_K)$

  - $y^*_K \Longleftarrow \arg\max_{y'_K} a_u(y'_K) + \alpha^*_{K-1}(y'_K)$
  - for $k = K$-1 to 1: $y^*_k \Longleftarrow \alpha^{\leftarrow}_k(y^*_{k+1})$

} backward decoding