TTIC 31230 Fundamentals of Deep Learning

Problems for Graphical Models.

Problem 1. Pseudolikelihood of a one dimensional spin glass. We let $\hat{x}$ be an assignment of a value to every node where the nodes are numbered from 1 to N_{nodes} and for every node i we have $\hat{x}[i] \in \{0,1\}$. We define the score of $\hat{x}$ by

$$f(\hat{x}) = \sum_{i=1}^{N-1} \mathbf{1}[\hat{x}[i] = \hat{x}[i+1]]$$

The probability distribution over assignments is defined by a softmax. We let $\hat{x}[i := v]$ be the assignment identical to $\hat{x}$ except that node i is assigned the value v. The expression $\hat{x}[i] = v$ is either true or false depending on whether no i is assigned value v in $\hat{x}$. So these are quite different.

$$P_f(\hat{x}) = \operatorname*{softmax}_{\hat{x}} f(\hat{x})$$

Pseudolikelihood is defined in terms of the softmax probability P_f as follows.

$$\tilde{P}_f(\hat{x}) \;\; = \;\; \prod_i P_f(\hat{x}[i] \mid \hat{x}\backslash i)$$

What is the **Pseudolikelihood** of the all ones assignment under the definition of f given above?

Solution:

$$\tilde{P}_f(\hat{x}) = \Pi_i P_f(\hat{x}[i] \mid \hat{x}/i)$$

where $\hat{x}/i$ consists of all components of $\hat{x}$ other than i. In a graphical model $P_f(\hat{x}[i] \mid \hat{x}/i)$ is determined by the neighbors of i and we can consider only how a value is scored against it neighbors. For $\hat{x}$ equal to all ones we have

$$
\begin{aligned}
f(\hat{x}) &= N-1 \\
f(\hat{x}[i := 0]) &= \left\{ \begin{array}{ll} N-3 & \text{for } 1 < i < N \\ N-2 & \text{for } i = 1 \text{ or } i = N \end{array} \right.
\end{aligned}
$$

For $1 < i < N$ we get

$$
\begin{aligned}
Q_f(\hat{x}[i = 1] \mid \hat{x}/i) &= \frac{e^{N-1}}{e^{N-1} + e^{N-3}} \\
&= \frac{1}{1 + e^{-2}}
\end{aligned}
$$

and for $i = 1$ or $i = N$ we get

$$Q_f(\hat{x}[i = 1] \mid \hat{x}/i) = \frac{1}{1 + e^{-1}}$$

This gives

$$\tilde{Q}(\hat{x}) = (1 + e^{-1})^{-2}(1 + e^{-2})^{-(N-2)}$$

Problem 2. Pseudolikelihood for images. Consider a semantic segmentation $\hat{y}[i]$ on pixels i with $\hat{y}[i]$ a semantic class label in $\{C_1, \ldots, C_K\}$. We also assume a scoring function s_Φ on semantic segmentations defining

$$P_\Phi(\hat{y}) = \operatorname*{softmax}_{\hat{y}} \; s_\Phi(\hat{y})$$

Pseudolikelihood is defined by

$$\tilde{P}_\Phi(\hat{y}) \;\;=\;\; \prod_i P_\Phi(\hat{y}[i] \mid \hat{y}\backslash i)$$

where $\hat{y}\backslash i$ assigns a class to every pixel other than i, and $\hat{y}[i := c]$ is the semantic segmentation identical to $\hat{y}$ except that pixel i is labeled with semantic class c. In a typical graphical model for images we have

$$P_\Phi(\hat{y}[i] \mid \hat{y}\backslash i) = P_\Phi(\hat{y}[i] \mid \hat{y}[N(i)])$$

where $\hat{y}[N(i)]$ is $\hat{y}$ restricted to those pixels which are neighbors of pixel i.

(a) show

$$\frac{P_\Phi(\hat{y})}{\sum_c P_\Phi(\hat{y}[i := c])} = \operatorname*{softmax}_c s_\Phi(\hat{y}[i := c]) \quad \text{evaluated at } c = y[i]$$

Solution:

$$\begin{aligned}
\frac{P_\Phi(\hat{y})}{\sum_c P_\Phi(\hat{y}[i := c])} &= \frac{\frac{1}{Z} e^{s_\Phi(\hat{y})}}{\sum_c \frac{1}{Z} e^{s_\Phi(\hat{y}[i:=c])}} \\
&= \frac{e^{s_\Phi(\hat{y})}}{\sum_c e^{s_\Phi(\hat{y}[i:=c])}} \\
&= \operatorname*{softmax}_c \; s_\Phi(\hat{y}[i := c]) \;\; \text{evaluated at } c = y[i]
\end{aligned}$$

(b) How many scores need to be computed in the worst case for computing $P_\Phi(\hat{y})$. Given the result of part (a), how many for computing $\tilde{P}_\Phi(\hat{y})$?

Solution: K^N for P_Φ and KN for $\tilde{P}_\Phi$.

(c) Consider a distribution on semantic segmentations where for each pixel the class assigned to that pixel is uniquely determined by the classes of its neighbors. Can this distribution be defined by a softmax over scores? Explain your answer.

Solution: No. Since $e^s > 0$ for any finite s, all semantic segmentations must have nonzero probability.

(d) If P_Φ is a distribution defined in some other way such that the class of each pixel is completely determined by the other pixels, given a simple expression for $\tilde{P}_\Phi(\hat{y})$ in the case where $\hat{y}$ has nonzero probability under P_Φ.

Solution: We have $P_\Phi(\hat{y}|\hat{y}\backslash i) = 1$ which implies $\tilde{P}(\hat{y}) = 1$.