Inference and Representation, Fall 2017

Structured Prediction for Part-of-Speech Tagging, MLE and Max-Ent, Max-Sum BP

Zhuoru Lin
zlin@nyu.edu

Disclaimer: I adhered to NYU honor code in this assignment.

1. **Part-of-speech tagging using SSVM**

2. **Max-Product**

3. **Exponential families**
The log-likelihood of the data is:

$$\log[p(x,\theta)] = \sum_{n=1}^{L}(\langle\theta, f(x^{(n)})\rangle - \log[Z(\theta)]) \tag{1}$$

The maximum likelihood estimation θ_{ML} must satisfied:

$$\frac{\partial p}{\partial \theta}(\theta_{ML}) = 0 \tag{2}$$

$$\sum_{n=1}^{L} f(x^{(n)}) - L\sum_{x}\frac{\exp[<\theta_{ML}, f(x^n)>]}{Z(\theta_{ML})}f(x^n) = 0 \tag{3}$$

$$\frac{1}{L}\sum_{n=1}^{L} f(x^{(n)}) = \sum_{x} p(x|\theta_{ML})f(x^n) \tag{4}$$

Technically, we also need to show that this is in fact an maximum by showing the secon-derivative. This is ignored here.

4. **Maximun entropy distribution** The maximum entropy distribution can be formulated as an functinal optimization problem as:

$$\begin{aligned} \arg\max_p \int_x \log(p(x))p(x)dx \\ \text{with} \int_x p(x)dx = 1 \\ \text{and} \int p(x)f_k(x) = \alpha_k \text{ for all k} \end{aligned}$$

The Lagrangian is:

$$\mathcal{L} = \int_x \log(p(x))p(x)dx - \sum_k (\theta_k \int p(x)f_k(x) - \alpha_k) - (\lambda \int p(x)dx - 1). \tag{5}$$

The Lagrange multiplier states that the optimum acheives zero derivatives of Lagragian. Taking the functional derivative, we have:

$$\log(p(x)) + 1 - \lambda - \sum_k \theta_k f_k(x) = 0 \tag{6}$$

Therefore we must have:

$$\begin{aligned} p(x) &= \exp(\sum_k \theta_k f_k(x) + \lambda - 1) && (7) \\ &= \frac{1}{Z} \exp(\langle \theta, f(x) \rangle) && (8) \end{aligned}$$

with $Z = \exp(1 - \lambda)$