

Weekly Exercises 9

Room: 02.09.023

Wednesday, 17.07.2019, 12:15 - 14:00

MRF/CRF learn (Due: 15.07) (6+6 Points)

NOTE: There will be another programming assignment next week on EM algorithm and graph cut (due 29.07).

Exercise 1 (6 Points). (Hessian of NLL) Prove the following result from the lecture (P9 of learning):

$$\nabla_{\theta}^2 l(\theta) = \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x)\psi(x)^{\top}] - \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x)] \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x)]^{\top} = \text{Cov}_{x \sim p(\cdot; \theta)}[\psi(x)] \quad (1)$$

Solution. We have that

$$\nabla_{\theta}^2 l(\theta) = \nabla_{\theta}^2 \log Z(\theta) \quad (2)$$

$$= J_{\theta} \left(\sum_x \frac{\exp(\theta^{\top} \psi(x))}{\sum_{x'} \exp(\theta^{\top} \psi(x'))} \psi(x) \right)^{\top} \quad (3)$$

$$= \sum_x \nabla_{\theta} \left(\frac{\exp(\theta^{\top} \psi(x))}{Z(\theta)} \right) \psi(x)^{\top} \quad (4)$$

$$= \sum_x \left(p(x; \theta) \psi(x) - \frac{\exp(\theta^{\top} \psi(x)) (\sum_x \psi(x) \exp(\theta^{\top} \psi(x)))}{Z^2(\theta)} \right) \psi(x)^{\top} \quad (5)$$

$$= \sum_x p(x; \theta) \psi(x) \psi(x)^{\top} - \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x)] \sum_x \left(\frac{\exp(\theta^{\top} \psi(x))}{Z(\theta)} \right) \psi(x)^{\top} \quad (6)$$

$$= \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x) \psi(x)^{\top}] - \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x)] \mathbb{E}_{x \sim p(\cdot; \theta)}[\psi(x)]^{\top} \quad (7)$$

$$= \text{Cov}_{x \sim p(\cdot; \theta)}[\psi(x)] \quad (8)$$

Exercise 2 (6 Points). (partially observed CRF) Denoting x_i, y_i, z_i as the conditioned / observed / hidden part of the i -th sample respectively, $r(y, x)$ the empirical data distribution and $r_x(x) = \sum_y r(y, x)$. Similar to the cases with fully observed CRF and partially observed MRF, the partially observed CRF has the following distribution:

$$p(y, z|x; \theta) = \frac{1}{Z(\theta; x)} \exp(\theta^{\top} \psi(x, y, z)), \quad (9)$$

and we learn the parameters θ by minimizing the negative conditional data likelihood

$$l(\theta) = -\frac{1}{|\mathcal{S}|} \sum_{1 \leq i \leq |\mathcal{S}|} \log p(y_i | x_i; \theta). \quad (10)$$

Prove that its gradient is

$$\nabla_{\theta} l(\theta) = -\mathbb{E}_{(x,y) \sim r} [\mathbb{E}_{z \sim p(\cdot | x, y; \theta)} [\psi(x, y, z)]] + \mathbb{E}_{x \sim r_x} [\mathbb{E}_{(y,z) \sim p(\cdot, \cdot | x; \theta)} [\psi(x, y, z)]] \quad (11)$$

Solution.

$$\nabla_{\theta} l(\theta) = -\frac{1}{|\mathcal{S}|} \sum_{1 \leq i \leq |\mathcal{S}|} \sum_x \sum_y \delta_{x=x_i} \delta_{y=y_i} \nabla_{\theta} \log \sum_z p(y, z | x; \theta) \quad (12)$$

$$= -\mathbb{E}_{(x,y) \sim r} [\nabla_{\theta} \log \sum_z p(y, z | x; \theta)] \quad (13)$$

$$= -\mathbb{E}_{(x,y) \sim r} \left[\frac{1}{p(y|x; \theta)} \sum_z \nabla_{\theta} p(y, z | x; \theta) \right] \quad (14)$$

$$= -\mathbb{E}_{(x,y) \sim r} \left[\frac{\sum_z \left(\frac{\psi(x,y,z) \exp(\theta^{\top} \psi(x,y,z))}{Z(\theta;x)} - \frac{\exp(\theta^{\top} \psi(x,y,z)) \sum_{y,z} \psi(x,y,z) \exp(\theta^{\top} \psi(x,y,z))}{Z^2(\theta;x)} \right)}{p(y|x; \theta)} \right] \quad (15)$$

$$= -\mathbb{E}_{(x,y) \sim r} \left[\frac{\sum_z (\psi(x, y, z) p(y, z | x; \theta) - p(y, z | x; \theta) \mathbb{E}_{(y,z) \sim p(\cdot, \cdot | x; \theta)} [\psi(x, y, z)])}{p(y|x; \theta)} \right] \quad (16)$$

$$= -\mathbb{E}_{(x,y) \sim r} \left[\sum_z \psi(x, y, z) p(z | y, x; \theta) - \mathbb{E}_{(y,z) \sim p(\cdot, \cdot | x; \theta)} [\psi(x, y, z)] \right] \quad (17)$$

$$= -\mathbb{E}_{(x,y) \sim r} [\mathbb{E}_{z \sim p(\cdot | x, y; \theta)} [\psi(x, y, z)]] + \mathbb{E}_{(x,y) \sim r} [\mathbb{E}_{(y,z) \sim p(\cdot, \cdot | x; \theta)} [\psi(x, y, z)]] \quad (18)$$

$$= -\mathbb{E}_{(x,y) \sim r} [\mathbb{E}_{z \sim p(\cdot | x, y; \theta)} [\psi(x, y, z)]] + \mathbb{E}_{x \sim r_x} [\mathbb{E}_{(y,z) \sim p(\cdot, \cdot | x; \theta)} [\psi(x, y, z)]] \quad (19)$$

Programming (Due:22.07) (12 Points)

In this programming exercise, you are asked to classify MNIST hand-written digits with Naive Bayes model and learn the parameters. See the ipython file for more details.