14/10/19

EE5301: Representation Learning

- $R_s(\theta) = R(\theta) + \lambda \sum_{m=1}^{M} KL(\text{Bern}(p) \| \bar{z}_m)$, where λ is a Lagrange multiplier — (1)

- Recall from class that the sparsity constraint does not depend on β. $\Rightarrow$ The update equations for β remain unchanged
- We only need to find the update equation for α_{ml}.

$$\frac{\partial R_s(\theta)}{\partial \alpha_{ml}} = \frac{\partial}{\partial \alpha_{ml}} \left[R(\theta) + \lambda \sum_{j=1}^{M} KL(\text{Bern}(p) \| \bar{z}_j) \right]$$

(from defn of $KL(p \| q)$)

$$= \frac{\partial R(\theta)}{\partial \alpha_{ml}} + \lambda \cdot \frac{\partial}{\partial \alpha_{ml}} \sum_{j=1}^{M} \left[p \cdot \log \frac{p}{\bar{z}_j} + (1-p) \cdot \log \frac{(1-p)}{1-\bar{z}_j} \right]$$

$$= \frac{\partial R(\theta)}{\partial \alpha_{ml}} + \lambda \frac{\partial}{\partial \alpha_{ml}} \left[p \log \frac{p}{\bar{z}_m} + (1-p) \cdot \log \frac{(1-p)}{(1-\bar{z}_m)} \right]$$

($\because$ other terms in the sum do not depend on α_{ml})

$$= \frac{\partial R(\theta)}{\partial \alpha_{ml}} + \lambda \cdot \left[-\frac{p}{\bar{z}_m} + \frac{(1-p)}{(1-\bar{z}_m)} \right] \cdot \frac{\partial \bar{z}_m}{\partial \alpha_{ml}} \quad \text{— (2)}$$

Recall: $\bar{z}_m = \frac{1}{N} \sum_{i=1}^{N} z_m^{(i)} = \frac{1}{N} \sum_{i=1}^{N} \sigma(\underline{\alpha}_m^T \cdot \underline{x}^{(i)} + \alpha_{m0})$

$$\Rightarrow \frac{\partial \bar{z}_m}{\partial \alpha_{ml}} = \frac{1}{N} \sum_{i=1}^{N} \sigma'(\underline{\alpha}_m^T \underline{x}^{(i)} + \alpha_{m0}) \cdot x_l^{(i)} \quad \text{— (3)}$$

[using MLP notation]

Plugging ③ in ② gives us the new expression for the partial derivative of the regularized cost function w.r.t. α_{ml}.

Use this new expression to update α_{ml} for the sparse AE problem.