

# COMS 4771 Spring 2017 Homework 5

Jin Tack Lim, jl4312

Problem 1.

(a)

- Initialization

random  $p^j(1), \dots, p^j(D)$  where  $j = 1, \dots, K$

$$\pi_j = \frac{1}{K}$$

- Expectation

The probability that n-th document belongs to the class i. For t-th iteration, the superscript t for each of  $\tau, \pi, p^j(1), \dots, p^j(D)$  are omitted.

$$\tau_{n,i} = \frac{\pi_i \prod_{d=1}^D p^i(d)^{x_n(d)}}{\sum_{j=1}^K \pi_j \prod_{d=1}^D p^j(d)^{x_n(d)}}$$

- Maximization Update  $p^i(1), \dots, p^i(D)$  and  $\pi_i$  where  $i = 1, \dots, K$

$$\pi_i^{(t+1)} = \frac{\sum_{q=1}^n \tau_{q,i}^{(t)}}{n}$$

$$p^{i(t+1)}(d) = \frac{\sum_{q=1}^n \tau_{i,q} x_q(d)}{\sum_{q=1}^n \tau_{i,q} l_q}$$

where  $d=1, \dots, D, q = 1, \dots, n$

(b)

The marginal distribution of  $x_n$  is

$$\frac{\hat{x}_n}{l_n}$$

This can be interpreted as the binomial distribution for each dimensions. So if  $l_n$  is very large and  $\frac{\hat{x}_n}{l_n}$  is very small, it can be approximated to the Poisson distribution, where  $\lambda = \hat{x}_n$ .

(c)

1. Similarity

- Both of them can reduce the dimensionality.

2. Difference

- PCA gives vectors in order such that the variance along the vector is high. (e.g. the given data has the highest variance along the first vector.) For (b), there's no guarantee.

- PCA gives vectors that are orthogonal to each other, since they are all eigenvectors. For (b), there's no guarantee.

(d)

Problem 2.

Problem 3.

We have  $q+2$  layers except the convolution layer. Input to each layer is  $a^p$  and output is  $z^p$  where  $p = 1, \dots, q+2$ .

$$R = \frac{1}{N} \sum_{n=1}^N \frac{1}{2} (y_n - z_n^{q+2})^2$$

1. the update equation for V

$$\begin{aligned} \frac{\partial R}{\partial V} &= \frac{\partial R}{\partial z^{q+2}} \frac{\partial z^{q+2}}{\partial a^{q+2}} \frac{\partial a^{q+2}}{\partial V} \\ &= -\frac{1}{N} \sum_{n=1}^N (y - z^{q+2}) z^{q+2} (1 - z^{q+2}) z^{q+1} \end{aligned}$$

Therefore, the update equation for V is

$$V^{t+1} = V^t - \eta \frac{\partial R}{\partial V}$$

2. the update equation for U

$$\begin{aligned} \frac{\partial R}{\partial U} &= \frac{\partial R}{\partial z^{q+1}} \frac{\partial z^{q+1}}{\partial a^{q+1}} \frac{\partial a^{q+1}}{\partial U} \\ &= \frac{\partial R}{\partial a^{q+2}} \frac{\partial a^{q+2}}{\partial z^{q+1}} \frac{\partial z^{q+1}}{\partial a^{q+1}} \frac{\partial a^{q+1}}{\partial U} \\ &= \left( -\frac{1}{N} \sum_{n=1}^N (y - z^{q+2}) z^{q+2} (1 - z^{q+2}) \right) V z^{q+1} (1 - z^{q+1}) z^q \end{aligned}$$

The term put in the bracket is the one we can reuse from the previous derivation. (Backprop)

Therefore, the update equation for U is

$$U^{t+1} = U^t - \eta \frac{\partial R}{\partial U}$$

## References

- [1]: <https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>
- [2]: <http://cs231n.github.io/convolutional-networks/>