ChO

67/100

Solutions Sheet

Nina Fischer and Yannick Zelle

December 13, 2021

Exercise 1

Let us consider z_n to be se from a previous previous iterration step. So we have :

$$J = \sum_n \sum_i z_n^i ||x_- \mu_i||^{\hat{A}^2}$$?

We will now show that this Loss function converges and therefore the k-means algorithm converges

Proof. **(a)** The reassiment step is give by

$$z_n^{i\,\prime} = \begin{cases} 1, & \text{if } i = argmin_j ||x_n - u_l||^2 \\ 0, & \text{otherwise} \end{cases}$$

So

$$\begin{aligned} J' &= \sum_n \sum_i z_n^{i\,\prime} ||x_n - u_i||^2 \\ &= \sum_n \min_i ||x_n - u_i||^2 \end{aligned}$$

So we have :

$$J' \leq J$$

not a complete proof −5P

Thus the e-step minimizes J

b For the reassigment of the mean we reassine the $\mu_i' s$ by definition with teh average of those x, which already minimizes J, so by construction if we reassign :

$$\mu_i' = \frac{\sum_n z_n^i x_n}{\sum_n z_n^i}$$

and

$$J'' = \sum_n \sum_i z_n^{i\prime} ||x_n - u_i'||^2$$

we have

$$0 \leq J'' \leq J' \leq J$$

this does not show that the M-step minimizes J −15P

(c) We thus conclude that the distortion measure or i.e. the loss function of the k-means algo converges towards 0 and thus at a given point there will be no more need to reassign the z_n^i or expressed differently k-means is guaranteed to converge. ✓

□ 10/30

Exercise 2

Exercise 3

Exercise 4

Exercise 2

- allow some tolerance when comparing the new and old labels -1P
- good job!

24/25

Exercise 4

- first plot is meaningless (rather plot a histogram)
- (a) correct
- (b) • use the new μ for calculating σ^2 -1P
  • the GMM is calculating the variance σ^2 but the norm() function expects the standard deviation σ → take the square root then your code works -1P

23/25

ES 9

Exo 3

We have

$$\mathcal{L}(\theta, \lambda) = \ell(\theta | D) + \lambda \left(\sum_k \pi_k - 1 \right)$$

We will now calculate the derivatives

$$\frac{\partial \mathcal{L}}{\partial \pi_i} = \frac{\partial}{\partial \pi_i} \ell(\theta | D) + \frac{\partial}{\partial \pi_i} \lambda \left(\sum_k \pi_k - 1 \right)$$

$$= \sum_n \frac{\partial}{\partial \pi_i} \log \sum_j \pi_j f_j(x_n | \theta_j) + \lambda$$

$$= \sum_n \frac{1}{\sum_j \pi_j f_j(x_n | \theta_j)} \frac{\partial}{\partial \pi_i} \sum_j \pi_j f_j(x_n | \theta_j) + \lambda$$

$$= \sum_n \frac{1}{\sum_j \pi_j f_j(x_n | \theta_j)} f_i(x_n | \theta_i) + \lambda$$

$$= \sum_n \frac{f_i(x_n | \theta_i)}{\sum_j \pi_j f_j(x_n | \theta_j)} + \lambda \quad ✓$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{\partial}{\partial \lambda} \ell(\theta | D) + \frac{\partial}{\partial \lambda} \lambda \left(\sum_k \pi_k - 1 \right)$$

$$= \sum_k \pi_k - 1 \quad ✓$$

b) By setting $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$ we have:

$$\sum_k \pi_k - 1 = 0 \Leftrightarrow \sum_k \pi_k = 1 \quad ✓$$

c) By setting $\frac{\partial \mathcal{L}}{\partial \pi_i} = 0$, we obtain:

missing -5P

d) missing -5P

10/20