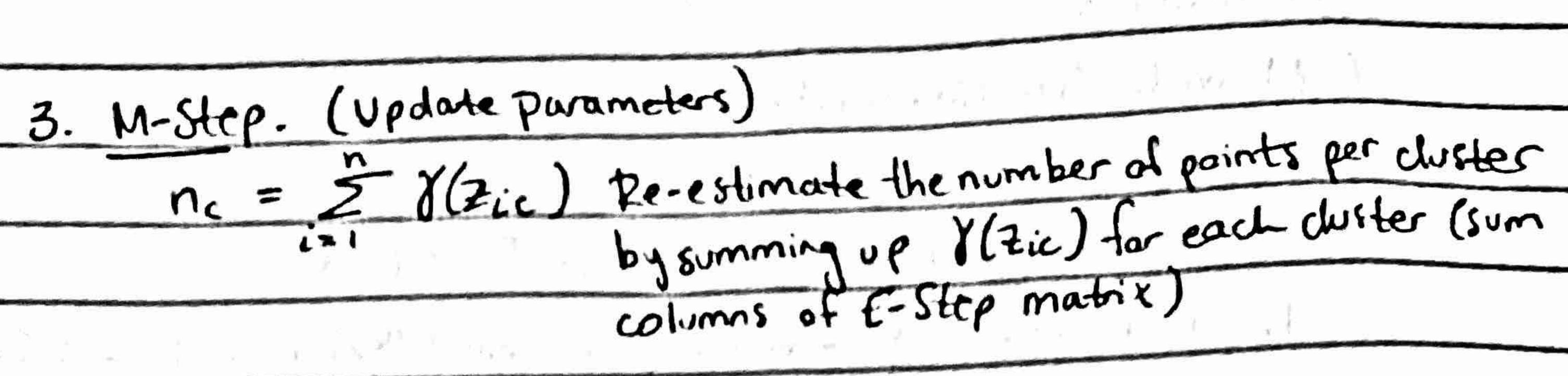
3/29/19 Discussion 8 E-M with Gaussian mixture models Expectation Maximization Algorithm for GMM Initialize IT = mixing weights (ne/total # of points) for each Gowssian p = mean for each Crowssian (random) Z = covariance matrix for each duster/Gaussian (Identity) he = number of points per cluster. This is usually set to be total # of points/k (H of chaters) initially. k = # of clusters (number of Gaussians to estimate) 2. E-Step For all i= 1 to n, c= 1 to k, compte: Y(2ic) = πc [2π d/2 |Zc|/2 txp (- ½(xi-νc) ] [-1 (xi-νc))] 0 = 1 [ = 1 (xi-mj) ] = (xi-mj) [ 2 (xi-mj) ] Probability that the ith point belongs to the ct duster. End result: an nxk matrix Sum of each row = 1, 0:1 0.4 .... Since probabilities that each 0.2 0.5 point lies in each duster will add to 1 (100%) In the above equation for O(Zic), I represents the dimensionality of the multivariate Gaussian (where d + k, the number of clusters, necessarily). If we have a GMM defined by 3 Gowssians, i.e.  $X = \{\chi_1, \chi_2, \chi_3\}$ , where  $\chi_i \sim \mathcal{N}(\omega, \Sigma)$ K = 3 = number of clusters/Gaussians, and d = number of dimensions in the data (how many teatures it has.)

7

P



$$M_c = \frac{nc/n}{M_c}$$

$$M_c = \frac{1}{nc} \frac{r}{r} \chi(z_{ic}) \chi_i$$

$$M_c = \frac{1}{nc} \frac{r}{i=1} \chi(z_{ic}) \chi_i$$

$$\sum_{c} = \frac{1}{n_c} \sum_{i=1}^{n_c} \mathcal{S}(z_{ic}) (x_i - \mu_c) (x_i - \mu_c)^T$$

4. Repeat until convergence.

the transfer of the second of the second of the second of

We are trying to maximize the log-likelihood estimate of our parameters. To check far convergence, we compute the log-likelihood and see if it is close enough to the previous estimate (less-than some convergence threshold).

$$MLE = \sum_{i=1}^{n} log \sum_{j=1}^{k} |T_{j}| \sum_{i=1}^{l} |I_{j}|^{1/2} exp(-\frac{1}{2}(x_{i}-\mu_{j}))^{T} \sum_{j=1}^{l} (x_{i}-\mu_{j})$$

This expression is the Mahalanobis

distance. In short, it measures the distance of a point to a distribution (in this case, the dister/

Gaussian). Minimizing this distance means the points are closer to the distribution (this is ideal,

Since we want our me and 2 tabe correct).

Due to the negative exponent, minimiting
the Mahalanobis distance will maximize
the MLE, which is what we want!