Haonan Hu

4/8/2021

2863545

1. Assuming a sample with 3 features, (6.0,3.0,5.0), is assigned by k- means clustering to the centroid, (7.3,4.3,6.0), what is the Euclidean distance from the centroid to the sample?

   d= $\sqrt[2]{(7.3-6)^2+(4.3-3)^2+(6-5)^2} = 2.09$

2. Assume a sample with 3 features, (6.0,3.0,5.0), is assigned by GMM to a component with a mean (μ) = (7.3,4.3,6.0) and a variance ($\sigma^2$) = (2.3,10.3,1.0) and that there is no covariance (i.e., all zeros, except along the diagonal of the covariance matrix), what is the Mahalanobis distance for the sample?

   D=$\sqrt[2]{(7.3-6)^2/2.3+(4.3-3)^2/10.3+(6-5)^2/1} = 1.38$

3. Assume you run GMM for k = 2 through 12 and get the following AIC scores: 57000, 52500, 45000, 41000, 39500, 38750, 38000, 37500, 37000, 36750, and 36500, respectively. Plot the AIC scores vs. k. Using the “elbow” method, what is the best number of GMM components for this application?

   Based on the method, **5** is the best number.