Questions A

A1. Definitions

- On slide 5, how is $\boldsymbol{\Sigma}$ defined?
- On slide 5, how is $\boldsymbol{\mu}$ defined? How can $\boldsymbol{\mu}$ be initialized with estimates?

A2. Covariance matrix I

- From the graph on slide 6, estimate values for g_k und $\boldsymbol{\mu}_k$.
- Which $\boldsymbol{\Sigma}_k$ are diagonal, which are not?

A3. EM algorithm I

- When using the EM algorithm, what problems might appear if only a few feature vectors are assigned to a specific class?
- How can these problems be avoided?

A4. EM algorithm II

- What is the meaning of the variable $\gamma(z_k(n))$?
- Why is the GMM sometimes called “multivariate density model”?

Questions B

B1. EM algorithm III

- Motivate that both equations for $p(\mathbf{X}|\mathbf{g},\boldsymbol{\mu},\boldsymbol{\Sigma})$ on slide 24 are equivalent.
- Across what index is the “outer” and the “inner” summation/multiplication applied?
- What is the effect of the denominator in the equations on slide 25?

B2. Covariance matrix II

- Which dimensions does $(\mathbf{x}(n)-\boldsymbol{\mu}_k^{\text{new}})$ have, which dimensions its transpose (slide 26, top)?
- Where can the variance of each feature dimension be found in the matrix $(\mathbf{x}(n)-\boldsymbol{\mu}_k^{\text{new}})\,(\mathbf{x}(n)-\boldsymbol{\mu}_k^{\text{new}})^T$?

B3. Latent random variables

- What is a latent random variable?
- For which purpose is it introduced?

B4. EM algorithm IV

- How can a GMM be initialized?

Answers B

B1. EM algorithm III

- The product sign can be put in front of the logarithm, and becomes a summation; Because the logarithm is monotonously increasing, the argmax is not being altered.
- The “inner” part, across *k*: Number of Gaussians.
  The “outer” part, across *n*: (Time-) index of the feature vectors.
- The denominator works as a normalization of the probabilities.

B2. Covariance matrix II

- (*D*x*1*) and (*1*x*D*), respectively. The result of the multiplication has the dimension (*D*x*D*), equivalent to the dimension of the covariance matrix.
- The variances of each feature dimension can be found at the diagonal of the *D*x*D*-matrix.

B3. Latent random variables

- A latent random variable is a “hidden” random variable, which influences the actual random variable and cannot be measured itself.
- Here, the latent random variable determines the assignment to a certain Gaussian, which cannot be measured.

B4. EM algorithm IV

- See slide 44.

Christian-Albrechts-Universität zu Kiel

Answers A

A1. Definitions

- $\boldsymbol{\Sigma}$ is defined by the covariances $\boldsymbol{\Sigma}_{i,j} = cov(x_i,x_j)$.
- $\boldsymbol{\mu}$ is defined as the mean value (i.e., the center) of the corresponding Gaussian. $\boldsymbol{\mu}$ can be initialized using a codebook.

A2. Covariance matrix I

- The averages $\boldsymbol{\mu}_k$ correspond to the coordinates of the maxima of the Gaussian curves, the weights g_k correspond to the volume ratios of the Gaussian curves (they sum up to 1).
- All $\boldsymbol{\Sigma}_k$ are diagonal, except $\boldsymbol{\Sigma}_k$ corresponding to the Gaussian at $\boldsymbol{\mu}_k$ = (-0,5; -0,5).

A3. EM algorithm I

- The Gaussian is likely to become a narrow peak (with low variance). See slides 28 and 29.
- Avoidance: Definition of a lower limit for the variance.

A4. EM algorithm II

- $\gamma(z_k(n))$ is the classification probability, a soft (or weighted) assignment of a feature vector to the k-th Gaussian distribution.
- The probability distribution of a multi-dimensional random variable is called *multi-dimensional* or *multivariate* distribution.