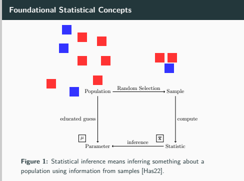

Population includes all of the elements from a set of data. Sample consists of one or more observations from the population.

Parameter Characteristic of a distribution describing a population, such as the mean or standard deviation of a normal distribution. Often notated using Greek letters.

Statistic A numerical value that represents a property of a random sample. Examples of statistics are

- the mean value of the sample data.
- the range of the sample data.
- deviation of the data from the sample mean.

Covariance Matrix

The covariance matrix of a multidimensonal variables is filled with individual variables, consider the two-dimensional case:

$$\Sigma = \begin{pmatrix} \hat{\sigma}_{xx} & \hat{\sigma}_{xy} \\ \hat{\sigma}_{yx} & \hat{\sigma}_{yy} \end{pmatrix} \tag{12}$$

$$\Sigma = (\mathbf{x}_i - \hat{\mu}_g)(\mathbf{x}_i - \hat{\mu}_g)^T \tag{13}$$

$$\Sigma = (\mathbf{x}_i - \hat{\mu}_g)(\mathbf{x}_i - \hat{\mu}_g)^T \tag{14}$$

$$= \begin{pmatrix} (\mathbf{x}_1 - \mu_1)^2 & (\mathbf{x}_1 - \mu_1)(\mathbf{x}_2 - \mu_2) & \dots & (\mathbf{x}_1 - \mu_1)(\mathbf{x}_n - \mu_n) \\ (\mathbf{x}_2 - \mu_2)(\mathbf{x}_1 - \mu_1) & (\mathbf{x}_2 - \mu_2)^2 & \dots & (\mathbf{x}_2 - \mu_2)(\mathbf{x}_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{x}_n - \mu_n)(\mathbf{x}_1 - \mu_1) & (\mathbf{x}_n - \mu_n)^2 & \dots & (\mathbf{x}_n - \mu_n)(\mathbf{x}_n - \mu_n) \end{pmatrix} \tag{15}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} \tag{16}$$

Gaussian mixture models

A Gaussian mixture model has the density [McN16]

$$f(\mathbf{x}|\theta) = \sum_{g=1}^{G} \rho_g \phi(\mathbf{x}|\mu_g, \Sigma_g). \tag{20}$$

With the normal distribution ϕ defined as before. ρ_g denotes the global probability with which a data value could originate from gaussian g. The gs number the gaussians, and G is the total number of Gaussians in the mix. We will use two. ϕ denotes the parameters μ_g and Σ_g.

Typically we want as many g as we have classes in the data. I.e. one for healthy and one for diabetic. The data vectors are p dimensional $\mathbf{x} \in \mathbb{R}^p$

Likelihood

Likelihood models the probability of data originating from a distribution as a function of the parameters. The gaussian case is modelled by [McN16]

$$\mathcal{L}_c(\theta) = \prod_{i=1}^{n}\prod_{g=1}^{G}[\rho_g \phi(\mathbf{x}_i|\mu_g, \Sigma_g)]^{z_{ig}}. \tag{21}$$

We want to maximize the likelihood.
In other words, we want to transform the bells in such a way, that they explain the points as plausible as possible.

To maximize *phi* it needs to sit on top of the points it labels. When a gaussian sits on top of many points it's ρ_g should be large. Finally, when this works well we want a big weight from z_{ig}.

Clustering using a GMM

After guessing an initial choice for all $\hat{\mu}_g$ and $\hat{\Sigma}_g$ [McN16],

$$\hat{z}_{ig} = \frac{\rho_g \phi(\mathbf{x_i}|\hat{\mu}_g, \hat{\Sigma}_g)}{\sum_{h=1}^{G} \rho_h \phi(\mathbf{x_i}|\hat{\mu}_h, \hat{\Sigma}_h)} \tag{23}$$

tells us the probability with which point x_i came from gaussian g. It creates an association between the data points and the Gaussians. Numerically evaluation results in a matrix $\mathbf{Z} \in \mathbb{R}^{G \times n}$. Use it's output to select the points which belong to each class.

The z_{ig} are the true labels, $\hat{z}_{ig}$ is our estimation. The $\hat{z}_{ig}$ are the expected value of the complete data log-likelihood. Why? ϕ is a pdf. A pdf can be interpreted as providing a relative likelihood that the value of the random variable would be equal to that sample [1]. We ask for all gaussians and every point and normalize.