[DATA SCIENCE]: EM algorithm:

- We might care about data density. (clusters?)
- If you know how the data is distributed, then you can do ANYTHING. Hence, why you need/assume gaussian distribution's do analysis.
- But the idea is that knowing the distribution means you know something about how the data/features related/contribute.
- [Mixture models:]
- mixture model → Family of functions w/ different parameters, in order to generalize and know something about the density of data.
- Formulaically, it is $f(x) = \sum_{j=1}^{c} \pi_j \phi(x, \theta_j)$.
  - where π_j is the mixture parameters, and $\phi(x, \theta_j)$ is the function controlled by the parameters θ_j.
- idea becomes that we've fit some functions, and then we think about how likely the data generates and fits our functions, if we sample some points in the space.
- [maximum log likelihood] (Should eval the formula for exam.),
- The idea is that we have many number of components, which help us better fit models to the data, in order to get the proper likelihood to get the sample that reflects the fitted model.
- 2-component mixture will give us a sum inside the log, which prevents us from what we need to do.
- Thus, [Expectation-Maximization].
- The idea is that this applies in cases where we don't know the labels (otherwise we would just calculate clusters.)
- The idea then, is to go about it as if we knew the labels, we have some idea on how [points x_i contributes] to the [density], (ie: how likely a point x_i contributes to a density parameter.)(this is the [expectation] step.)
- EM is a type of soft assignment, where we have a sum of the two (or more) distributions.

this is the core concept:

θ^0 generates γ_i for every i, and that generates θ_1, and guess, process repeats, and we hope to converge.

(It's a type of hill climbing, sensitive to local maxima)