

Discussion 3

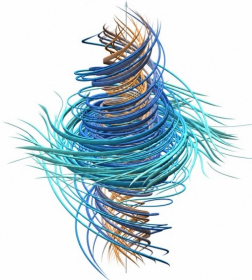


fig: A proposed vortex that
solves the Navier-Stokes
problem

This week's resources:

- 3Blue1Brown on Bayes Theorem [YouTube video](#)
- Steve Brunton: MLE with Examples [YouTube video](#)
- OpenAI found a solution to a Millenium Prize Problem with an AI model! [link](#)

1 Gaussian Mixture Clustering with Kangaroos and Berkeley Students (1D, Two Gaussians)

Consider a Gaussian Mixture Model (GMM) with two components in one dimension. Let $z_n \in \{K, B\}$ be the latent (unobserved) class of the n -th jumper (K = Kangaroo and B = Berkeley student) and $x_n \in \mathbb{R}$ be the observed jump height (in meters).

Assume that the classes occur with equal probability and that, conditional on the class, jump heights are Gaussian with class-dependent means μ_K and μ_B (the average jump heights of Kangaroos and Berkeley Students, respectively), and a shared variance σ^2 (the variability in jump heights).

(a) Write down the following quantities explicitly.

- The **prior** probability that a jumper is a Kangaroo or Berkeley student before observing the jump height: $p(z_n)$.
- The **likelihood** of the jump height conditioned on whether the jumper is a Kangaroo or Berkeley student: $p(x_n | z_n)$.

- (iii) The the **marginal** probability that a certain jump height is observed: $p(x_n)$.

- (iv) The **posterior** probability of a jumper being a Kangaroo or Berkeley student after observing the jump height: $p(z_n \mid x_n)$.

(b) Recall the K-means objective:

$$J(\{r_{nk}\}, \{\mu_k\}) = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2$$

where μ_k is the mean of cluster k and $r_{nk} \in \{0, 1\}$ is a binary indicator variable that describes which of the K clusters the data point x_n is assigned to. The K-means algorithm alternates between updating the cluster assignments and updating the cluster centers.

We modify this algorithm to generate GMMs in the following way:

- (1) **Update assignments:** Replace the hard assignments r_{nk} with soft assignments γ_{nk} , representing the probability that a point x_n belongs to cluster k .
- (2) **Update parameters:** Rather than selecting cluster centers that minimize the squared distances of data points to their assigned cluster centers, select the parameters θ that maximize the Q function:

$$Q(\theta) = \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} \ln p(x_n, z_n = k \mid \theta).$$

- (i) Identify which probability from part (a) corresponds to γ_{nk} . Verify $\sum_k \gamma_{nk} = 1$.

- (ii) Show that the Q function is actually the expected log-likelihood of our observed data, $X = (x_1, \dots, x_N)$, and the (unobserved) latent vector, $Z = (z_1, \dots, z_N)$ under the posterior probability of Z .¹

- (iii) Derive the maximum likelihood estimate for μ_k using the *expected* log-likelihood, $Q(\theta)$, treating γ_{nk} as constants.

¹We optimize the joint likelihood because it is much more tractable to compute while still optimizing the marginal likelihood. See this page for more information.

2 Binomial MLE with Misclassification

(a) You run m independent experiments. In each experiment, you perform n independent Bernoulli trials with unknown true success probability p . However, your measurement device is imperfect:

- A true success is recorded as a success with probability $1 - q$ (and as a failure with probability q).
- A true failure is recorded as a success with probability r (and as a failure with probability $1 - r$).

Let $X_i \sim \text{Bin}(n, p)$ denote the unobserved number of true successes in experiment i , and let Y_i be the observed number of recorded successes. Recall that the probability mass function for the binomial random variable $X_i \sim \text{Bin}(n, p)$ is

$$p(x_i) = \binom{n}{x_i} p^{x_i} (1 - p)^{n - x_i}.$$

Derive the likelihood for the observations $\mathcal{D} = \{y_i\}_{i=1}^m$. *Hint: Your answer should be a function of the true success probability p .*

(b) Find the MLE $\hat{p}$ in closed form. Show all steps.

