Proposal

Probabilistic Machine Learning Methods In Functional Mixture Modelling (FMM)

Ahsan Mir
25100325@lums.edu.pk

Lahore University of Management Sciences
Bachelor of Science in Computer Science

1 Introduction

In the ML setup of the original deterministic implementation, the relationship between impedance, Z, and concentration, C, is represented in the following fitted logarithmic model,

$$\mathbf{Z} = \alpha(\omega)\log(C) + \beta(\omega)$$

Where:

- $\mathbf{Z} \in \mathbb{R}^{27}$ is the vector of impedance values measured at 27 different frequencies in the range 100 Hz - 50 KHz.
- $\alpha(\omega) \in \mathbb{R}^{27}$ and $\beta(\omega) \in \mathbb{R}^{27}$ are the frequency-dependent parameter vectors with each element corresponding to a different frequency, ω, in the range 100 Hz - 50 KHz.
- $C \in \mathbb{R}$ is the concentration of yeast, which we are trying to estimate.

The objective is to use the fitted model to generate impedance data at known concentrations (to simulate the ground truth), add realistic noise to the generated impedance values, and test how well the regression algorithm can recover the true/original concentration from them. The goal of this proposal is to introduce a Bayesian perspective for modeling our parameters probabilistically, allowing us to obtain a more comprehensive understanding of the uncertainty in the model's predictions.

2 Objective

The initial approach provided a deterministic way to obtain concentration estimates by assuming fixed point estimates for the parameters α and β without considering the uncertainty in these estimates. This limitation motivates the shift to a probabilistic machine learning approach to explicitly model and quantify the uncertainty inherent in the parameters of our model. We propose treating these parameters as **random variables**, allowing us to propagate uncertainty through the model:

$$\alpha_i \sim \mathcal{N}(\mu_{\alpha_i}, \sigma^2_{\alpha_i}), \quad \beta_i \sim \mathcal{N}(\mu_{\beta_i}, \sigma^2_{\beta_i}), \quad i = 1, 2, \ldots, 27$$

Where:

- $\mu_{\alpha_i}, \mu_{\beta_i}$ are the means of the distributions of α_i and β_i, respectively.
- $\sigma^2_{\alpha_i}, \sigma^2_{\beta_i}$ are the variances, representing the **uncertainty** in the parameters.

Now, α and β are no longer fixed but have associated distributions that account for our confidence in their values.

3 Methodology

3.1 Posterior Inference on α and β

3.1.1 Prior Distributions for α and β

To define the prior beliefs about our parameters, we assign the prior distributions as follows:

$$\alpha_i \sim \mathcal{N}(\mu_{\alpha_i}, \sigma^2_{\alpha_i}), \quad \beta_i \sim \mathcal{N}(\mu_{\beta_i}, \sigma^2_{\beta_i}), \quad i = 1, \ldots, 27$$

where μ_{α_i} and μ_{β_i} are obtained from the fitted model weights. However, σ_{α_i} and σ_{β_i} are not directly provided by the fitted model. Instead, we can estimate these variances using the Laplace approximation to approximate the posterior as a Gaussian distribution.

3.1.2 Likelihood Function For True Impedance Values

From the fitted logarithmic model, we obtain a noiseless relationship between concentration and impedance with the fitted weights α_{true} and β_{true}. For a given true concentration, $C_{\text{true}} \in \mathbb{R}$, we can use this relationship to obtain the corresponding true impedance values, $Z_{\text{true}} \in \mathbb{R}^{27}$, at different frequencies,

$$Z_{\text{true},i} = \alpha_{\text{true},i} \cdot \log(C_{\text{true}}) + \beta_{\text{true},i}, \quad i = 1, 2, \ldots, 27$$

Since our observed impedance data is affected by measurement noise, the measured impedance values, Z_{measured}, differ slightly from the true impedance values, Z_{true}. We assume this measurement noise is Gaussian with mean zero and variance $\sigma^2_{\epsilon_i}$. The original experimental setup already estimates the mean noise level for each frequency by analyzing the standard deviation of the experimental impedance measurements across different replicates for the same concentration. Treating the mean noise level at the i-th frequency as the standard deviation of the Gaussian noise, $\sigma^2_{\epsilon_i}$ would simply be the square of this estimation.

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\epsilon_i}), \quad i = 1, 2, \ldots, 27$$

This means the noisy observed impedance, $Z_{\text{measured},i}$, can be modeled as:

$$Z_{\text{measured},i} = Z_{\text{true},i} + \epsilon_i = \alpha_i \cdot \log(C_{\text{true}}) + \beta_i + \epsilon_i, \quad i = 1, 2, \ldots, 27$$

We can now define our likelihood function as a product of 27 Gaussian distributions (one for each frequency) to represent the probability of observing the given set of measured impedance values, Z_{measured}, conditioned on the parameters α, β, and the true concentration C_{true}:

$$p(Z_{\text{true}} \mid \alpha, \beta, C_{\text{true}}) = \prod_{i=1}^{27} \mathcal{N}\left(Z_{\text{true},i} \mid \alpha_i \log(C_{\text{true}}) + \beta_i, \sigma^2_{\epsilon}\right)$$

Each Gaussian distribution describes how likely it is to observe an impedance value $Z_{\text{true},i}$, given the predicted value from our model $\alpha_i \log(C_{\text{true}}) + \beta_i$, and a noise variance σ^2_{ϵ}.

3.1.3 Posterior Distribution

At this point, we have defined the prior distributions for our parameters as well as the likelihood of observing the data for given parameter values. According to Bayes' Theorem, our posterior distribution, which represents the updated belief about the parameters, follows the relationship defined below:

$$p(\alpha, \beta \mid Z_{\text{true}}, C_{\text{true}}) \propto p(Z_{\text{true}} \mid \alpha, \beta, C_{\text{true}}) \cdot p(\alpha) \cdot p(\beta)$$

Where:

- $p(Z_{\text{true}} \mid \alpha, \beta, C_{\text{true}})$ is the likelihood function.
- $p(\alpha)$ and $p(\beta)$ are the prior distributions of the parameters.

However, we do not have a direct way to obtain a closed-form expression for the posterior using this relationship. Instead, we can derive an approximation for our posterior using the Maximum A Posteriori (MAP) estimate and the Laplace approximation.

3.1.4 Maximum A Posteriori (MAP) Estimation

Using the Bayes' theorem, we have formulated the following relationship:

$$p(\alpha, \beta \mid Z_{\text{true}}, C_{\text{true}}) \propto p(Z_{\text{true}} \mid \alpha, \beta, C_{\text{true}}) \cdot p(\alpha) \cdot p(\beta)$$

Since the log function is monotonic and does not change the location of the maximum, we can take the logarithm to simplify the optimization problem:

$$\mathcal{L}(\alpha, \beta) = \log p(\alpha, \beta \mid Z_{\text{true}}, C_{\text{true}}) = \log p(Z_{\text{true}} \mid \alpha, \beta, C_{\text{true}}) + \log p(\alpha) + \log p(\beta) + \text{constant}$$

To find the MAP estimates for our parameters (mode/peak of the posterior), we need to solve the optimization problem of maximizing this log-posterior. This can be done using various optimization algorithms like gradient ascent or Newton's method. The MAP estimates are given by:

$$(\alpha_{\text{MAP}}, \beta_{\text{MAP}}) = \arg\max_{\alpha,\beta} \mathcal{L}(\alpha, \beta)$$

3.1.5 Laplace Approximation

Once we have the MAP estimate, we can use the Laplace approximation to approximate the posterior around the MAP estimate. The Laplace approximation approximates the posterior as a Gaussian distribution centered at the MAP estimate, with a covariance matrix determined by the Hessian of the negative log-posterior at the MAP point.

The Hessian matrix $\nabla^2 \mathcal{L}(\alpha, \beta)$ is a matrix of second-order partial derivatives of the log-posterior:

$$\nabla^2 \mathcal{L}(\alpha, \beta) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \alpha} & \frac{\partial^2 \mathcal{L}}{\partial \beta^2} \end{bmatrix}$$

This matrix tells us about the curvature of the log-posterior around the MAP estimate. A steeper curvature indicates lower uncertainty, and vice versa. Using the Hessian of the negative log-posterior evaluated at the MAP estimate, we can obtain the covariance matrix of the Gaussian, Σ:

$$\Sigma = \left(-\nabla^2 \mathcal{L}(\alpha, \beta) \mid_{\alpha_{\text{MAP}}, \beta_{\text{MAP}}}\right)^{-1}$$

Finally, the posterior $p(\alpha, \beta \mid Z_{\text{true}}, C_{\text{true}})$ is approximated as:

$$p(\alpha, \beta \mid Z_{\text{true}}, C_{\text{true}}) \approx \mathcal{N}((\alpha_{\text{MAP}}, \beta_{\text{MAP}}), \Sigma)$$

This posterior approximation gives us a distribution over the parameters rather than just a single point estimate. This means instead of saying "the parameter α_i is equal to 2.5," we can now say, "given the observed data, the parameter α_i is likely to lie within a certain range, and this is how probable each value is." This helps us to capture the uncertainty about the true values of α and β. For example, if the data is noisy or limited, the posterior will reflect that uncertainty by having a wider spread.

3.2 Generating Noisy Impedances with K Samples

We now generate K noisy impedance vectors corresponding to our true concentration, C_{true}, to obtain our noised impedance matrix, $Z_{\text{noisy}} \in \mathbb{R}^{27 \times K}$,

$$Z_{\text{noisy}}^{(k)} = Z_{\text{true}} + \epsilon^{(k)}, \quad \epsilon^{(k)} \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad k = 1, 2, \ldots, K$$

The rationale behind choosing $K > 1$ noisy vectors instead of $K = 1$ is that we want to have a richer set of data representing the variability of impedance measurements at C_{true} to approximate the likelihood and perform Bayesian inference more accurately. With K vectors, each would provide an independent estimate of the posterior distribution for the parameters, resulting in a more informative posterior and stronger Bayesian update. Furthermore, instead of a single Monte Carlo (MC) integration, aggregation across both K noisy vectors and multiple posterior samples of the parameters can lead to a significant reduction in variance for our prediction of C_{true}. This would also potentially lead to tighter confidence intervals.

3.3 Posterior Sampling

Using the approximated posterior of α and β, we sample N sets of parameter values:

$$(\alpha_i, \beta_i) \sim \mathcal{N}((\alpha_{\text{MAP}}, \beta_{\text{MAP}}), \Sigma), \quad i = 1, 2, \ldots, N$$

3.4 Monte Carlo Estimation For Concentration Prediction

For each of the N samples from the posterior, $(\alpha^{(i)}, \beta^{(i)})$, and for each of the K noisy impedance vectors $Z^{(k)}_{\text{noisy}}$, we can now compute a concentration estimate $\mathcal{C}^{(i,k)}$ for each noisy impedance vector using a sampled set of parameters:

$$\mathcal{C}^{(i,k)}(Z^{(k)}_{\text{noisy}} \mid \alpha^{(i)}, \beta^{(i)}) = \exp\left(\frac{Z^{(k)}_{\text{noisy},j} - \beta^{(i)}_j}{\alpha^{(i)}_j}\right), \quad j = 1, \ldots, 27$$

We then average over all K noisy vectors and N posterior samples:

$$\hat{C}_{\text{final}} = \frac{1}{K \times N} \sum_{i=1}^{N} \sum_{k=1}^{K} \mathcal{C}^{(i,k)}$$

This provides an overall estimate of the concentration by averaging across multiple realizations of the noise and different parameter samples.
Once we have computed a concentration estimate for each noisy impedance realization and posterior parameter sample, we can simply aggregate to obtain the final concentration prediction:

$$\hat{C}_{\text{aggregated}} = \text{GeoMean}(\hat{C}_{\text{final}})$$

Since the original data and fitting were in the log space, using the geometric mean is suitable here to make the aggregation more consistent.

3.5 Error Estimation and Uncertainty Quantification

To quantify the uncertainty, we compute the variance over the K predictions for each of the N posterior samples:

$$\text{Var}(\hat{C}) = \frac{1}{K} \times \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \left(\mathcal{C}^{(i,k)} - \hat{C}_{\text{aggregated}}\right)^2$$

This variance captures the uncertainty introduced by both the noise in the impedance measurements and the uncertainty in the parameter estimates. The 95% confidence interval for our prediction can be computed as:

$$\hat{C}_{\text{aggregated}} \pm 1.96\sqrt{\text{Var}(\hat{C})}$$

This interval helps us understand how confident we are in the predicted concentration value.