EE5311 Differentiable and Probabilistic Computing: Assignment 1

XU NUO, A0313771H

1. Idea

1.1 Polynomial Regression Model

After observing the data trends, the simplest model that comes to mind is a polynomial regression model. I first used a second-order polynomial regression to describe the changes in platelet levels, with the mathematical model as follows:

$$P(t, A, B) = w_0 + w_1 t + w_2 t^2 + w_3 A + w_4 B + w_5 AB + w_6 A^2 + w_7 B^2 + w_8 tA + w_9 tB + \epsilon$$

where w_i represents the regression coefficients, which are fitted using the least squares method, and ϵ is the error term. This model considers linear terms, quadratic terms, and interaction terms (to account for mutual effects). The mathematical objective of regression fitting is as follows:

$$\min_{w_0, w_1, \ldots, w_9} \sum_{i=1}^{N} \left(P_i - \hat{P}(t_i, A_i, B_i) \right)^2$$

where P_i represents the observed experimental data, and $\hat{P}(t_i, A_i, B_i)$ is the predicted value from the model.

1.2 Logistic + Drug Effect Model

Inspired by the "sustainable fishing problem" mentioned in class, I developed a model based on logistic growth and incorporated the effects of drugs. The model assumes that platelet growth has an upper limit, represented by a maximum capacity K, and that the growth rate slows down as the concentration increases. Drugs A and B influence platelet levels, with their effects decaying over time. The natural decay of platelets is controlled by the decay coefficient γ. The final mathematical model is as follows:

$$\frac{dP}{dt} = rP(K - P) + \alpha A e^{-\lambda t} + \beta B e^{-\mu t} - \gamma P$$

Here, r represents the platelet growth rate, and K denotes the saturation level of platelets. The coefficients α and β describe the effects of drugs A and B on platelet levels, respectively, while A and B represent the dosages of drugs A and B. The parameters λ and μ control the time-dependent decay of the drugs, and γ governs the natural decay of platelets.

1.3 Physics-Informed Neural Networks

This method employs a neural network to learn the ODE using the PINN approach. The core idea is: The inputs include time t, the current platelet level P, drug dosages A and B, the output is the rate of change of platelet level $\frac{dP}{dt}$ and a fully connected neural network with 4 layers and *Tanh* activation functions is used to approximate the ODE. The mathematical model is as follows:

$$\frac{dP}{dt} = f(t, A, B, P)$$

where $f(t, A, B, P)$ is an unknown dynamic function learned by the neural network. The loss function consists of 2 parts: data loss and physics loss. The total loss function is:

$$\mathcal{L} = \sum \left(P_{pred} - P_{true}\right)^2 + \lambda_{phy} \sum \left(\frac{dP_{pred}}{dt} - f\left(t, A, B, P_{pred}\right)\right)^2$$

where λ_{phy} controls the impact of the physics constraint.

2. Results and Analysis

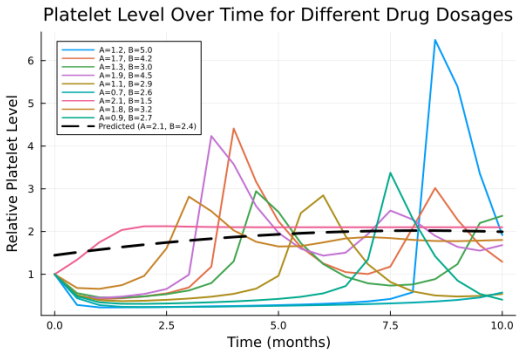

Fig. 1. Prediction using polynomial regression

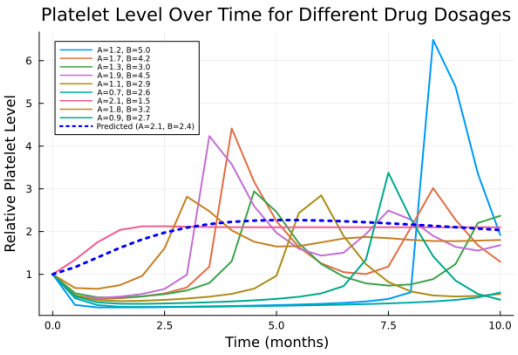

Fig. 2. Prediction using Logistic + Drug Effect

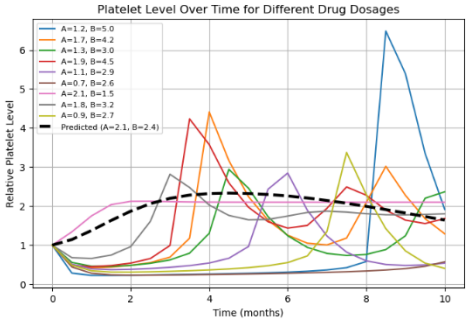

Fig. 3. Prediction using PINN

From the three figures, it can be seen that the predictions from different models are relatively similar, with smooth curves and platelet predictions within a normal range. However, in Figure 1, the predicted curve does not start from (0,1), showing a deviation from the actual statistical data. This indicates that polynomial fitting without any constraints still has flaws.

None of the three models produced the expected single or double peak patterns as in the existing data, which means there is a discrepancy between the predictions and the observed data trends. I have tried multiple approaches, including pharmacokinetic models and pure neural network fitting, but none have yielded entirely satisfactory results. I suspect that the issue lies in the mathematical modeling approach, where an inaccurate understanding of the quantitative relationships between variables has led to suboptimal predictions. Further exploration and research are needed.