Personalized medicine

Devising Chemotherapeutic Treatment regimen for Vincristine

Rohan Kumar

July 19, 2016

IIT Bombay

Table of contents

Introduction

Problem Recap

- Vincristine (VCR):
    - Chemotherapeutic drug, ALL, NHL
    - Neural and Non-Neural side effects
- Vincristine-induced Peripheral Neuropathy (VIPN)
    - Dose limiting neural side effect
    - Predict by analysis of metabolic profile
- Pharmacometabolomics approach
    - Identify and model biomarkers (metabolites)
    - Feature selection using Machine Learning algorithms

Feature Selection

Univariate Analysis

Determines strength of the relationship of a feature with the response variable.

- Pearson product-moment Correlation Coefficient
- Distance Correlation Coefficient
- Maximal Information Coefficient (MIC)

- Linear correlation
- Value in $[-1, 1]$
  - $-1/+1$: perfect linear correlation
  - 0 : no **linear** correlation

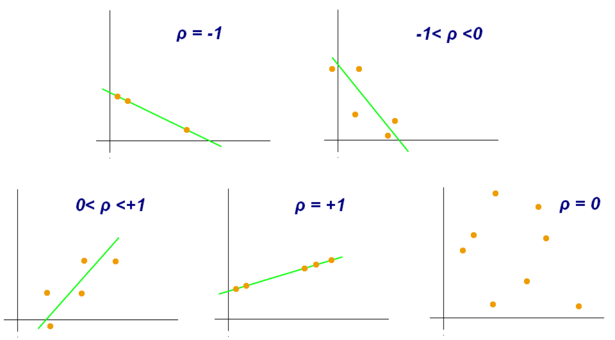


Figure 1: Scatter diagrams with different values of correlation coefficient

Image Source: Wikipedia

Univariate Analysis - Distance Correlation and MIC

- **Distance Correlation:** Correlation value of 0 implies independence.
- **Maximal Information Coefficient:** Strength of linear or non-linear association between X and Y

Univariate Analysis - Feature Selection

58 features selected.

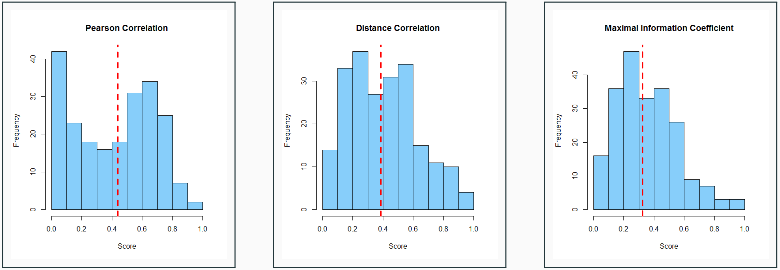


Figure 2: Histogram Plots for Univariate Analysis

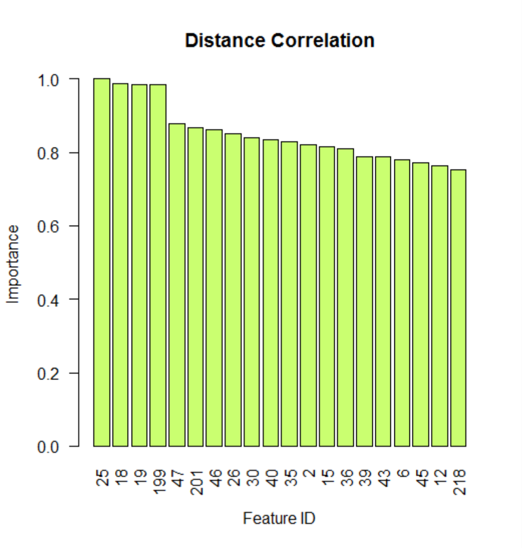


Figure 3: Feature Selection by Distance Correlation

Univariate Analysis - Feature Selection

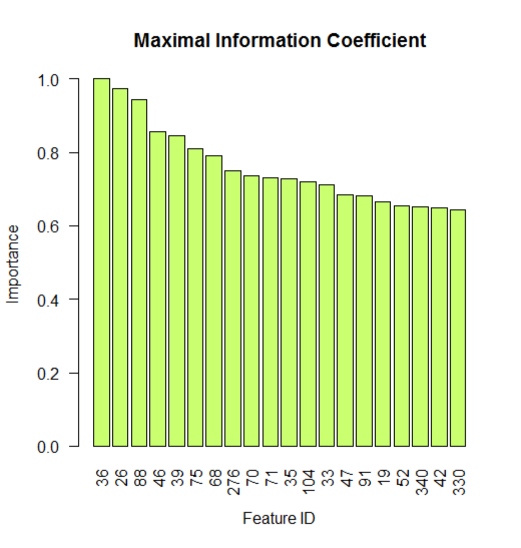


Figure 4: Feature Selection by MIC

Regularization

For the system

$$y = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

Linear regression solves the problem

$$minimize \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

In regularization we solve the problem

$$minimize \left(\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} (\beta_j)^2 \right)$$

Regularization

Penalty factor: $(\,\alpha\lambda_1 + (1-\alpha)\,\lambda_2\,)$

- **Lasso regression:** $\alpha = 1$
- **Ridge regression:** $\alpha = 0$
- **Elastic net regression:** $\alpha\,\epsilon\,(0, 1)$

Introduces bias in the method

Regularization - Feature Selection

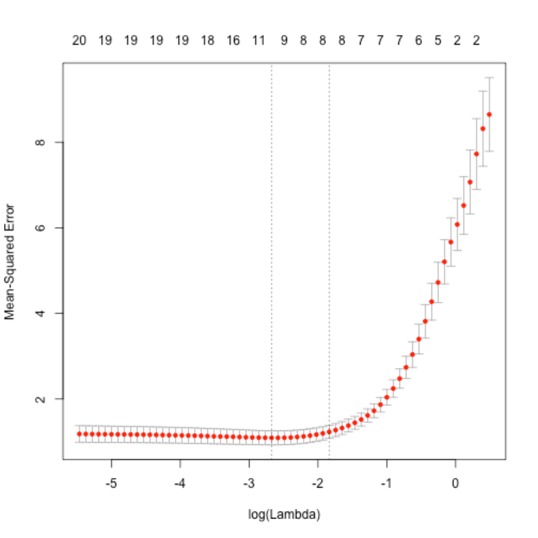


Figure 5: Selecting λ

Regularization - Feature Selection

- Run 100 regressions on dataset, note the frequency of appearance of features in successful runs
- Successful run: Area under Receiver Operating Characteristic (ROC) curve > 0.6
- Feature selected if $\beta_j \neq 0$
- Normalized to a range of $[0, 1]$ - Importance

Regularization - Feature Selection

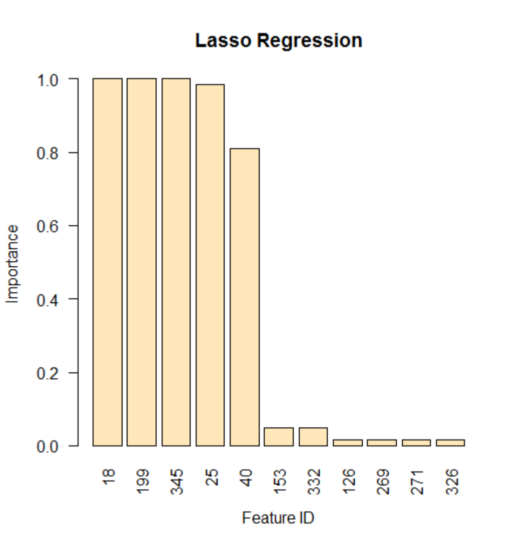


Figure 6: Feature Selection by Lasso Regression

Regularization - Feature Selection

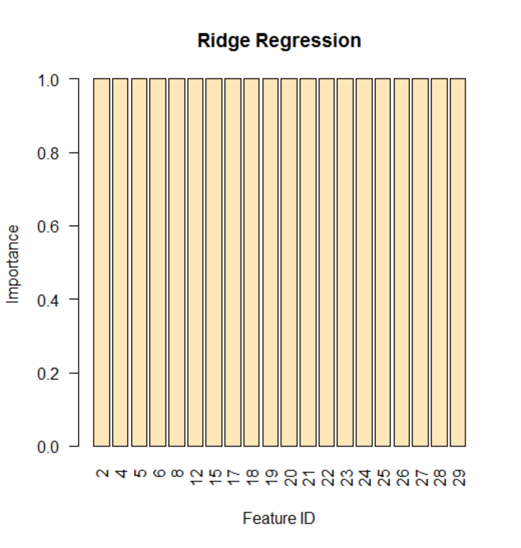


Figure 7: Feature Selection by Ridge Regression

Regularization - Feature Selection

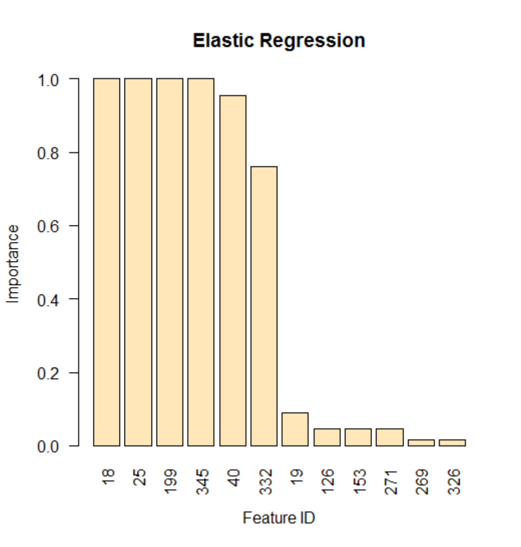


Figure 8: Feature Selection by Elastic Net Regression

A decision tree is a decision support tool that uses a tree-like graph.

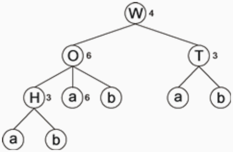


Figure 9: Sample Decision tree

Bagging, Random Forests, and Boosting use trees to construct more powerful prediction models.

Image Source: Wikipedia

- Build decision trees on bootstrapped training samples
- Randomly choose m predictors from p predictors at splits
- Train data AUC $= 1$
- Gini Index

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

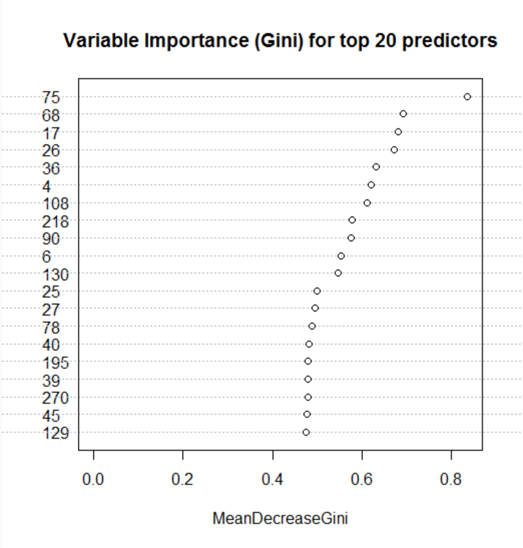
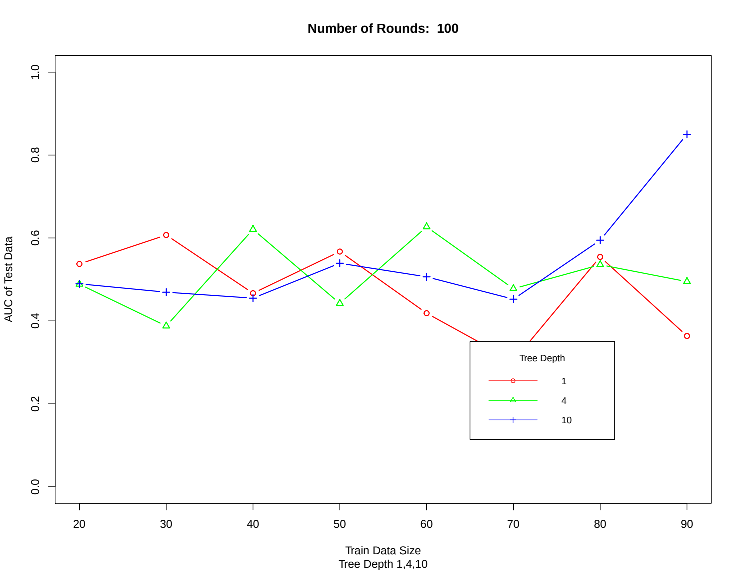


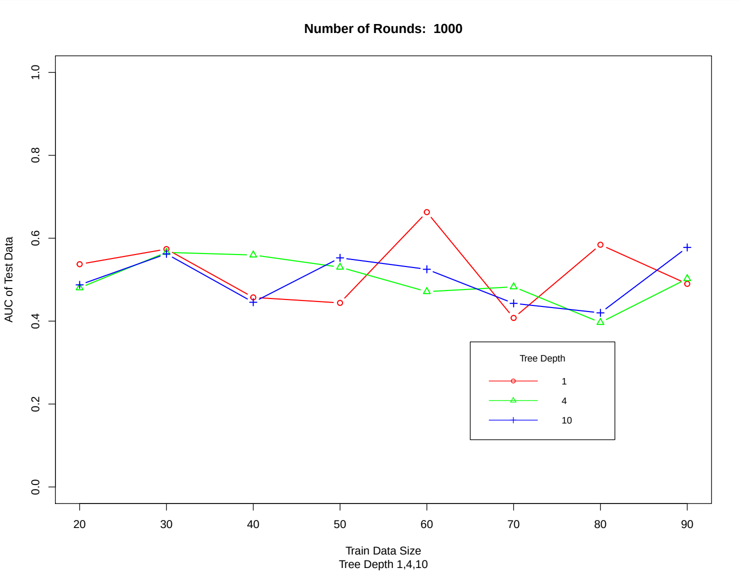
Figure 10: Feature Selection by Random Forest

- Fit a decision tree to a copy of the original training data set
- Repeat using information from the previously grown tree(s)
- Combine - single predictive model.
- Function *xgb.importance* in library *xgboost* in R

Decision Trees- Deciding tree depth and number of rounds



Decision Trees- Deciding tree depth and number of rounds



Decision Trees - Feature Selection

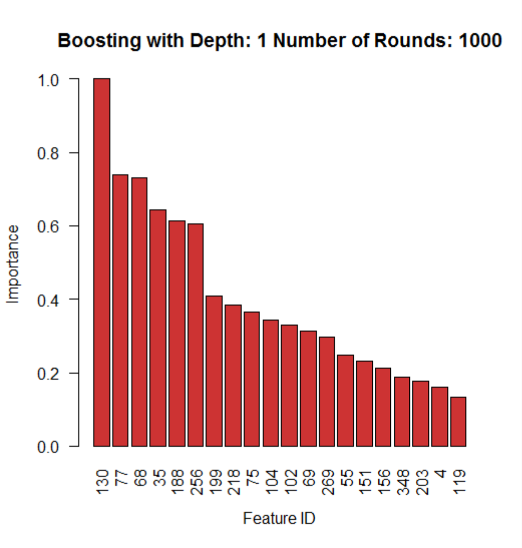


Figure 11: Feature Selection by boosted Decision Trees

Final Feature Selection

- Importance rating to each algorithm in $[0, 1]$
- Importance value from above algorithms
- Displays final features
- Model on Boosted Decision Trees

Final Feature Selection

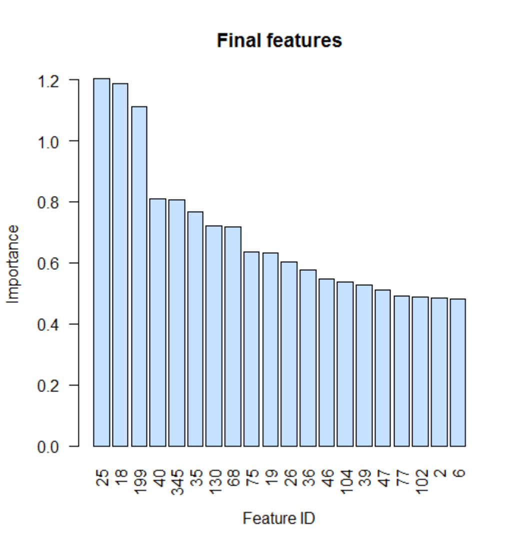
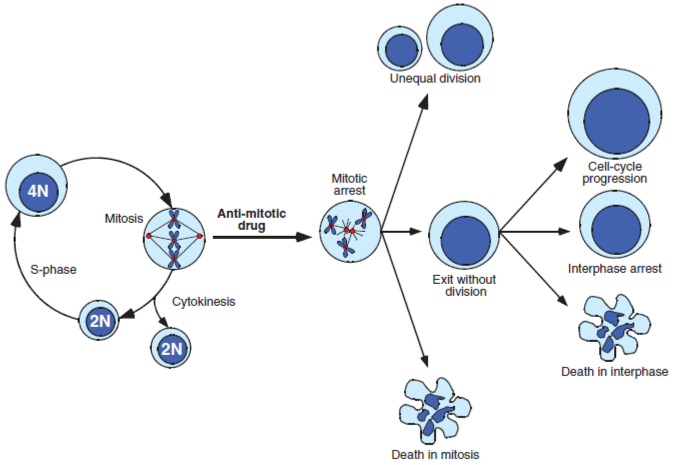


Figure 12: Final Features Selected

Timing drug dosage to increase efficacy

Pharmacodynamic Model for anti-mitotic drug



How do anti-mitotic drugs kill cancer cells?

Pharmacodynamic Model for anti-mitotic drug

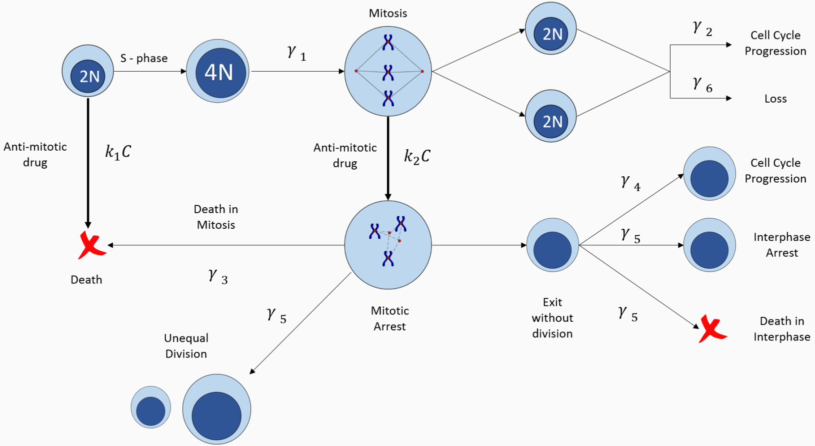
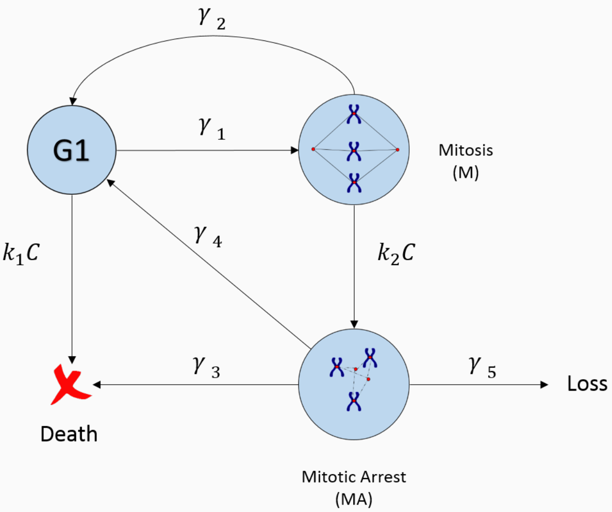


Image Source: Pharmacodynamic Model for anti-mitotic drug



Population balance

$$\frac{\partial n_{G1}}{\partial t} + \frac{\partial n_{G1}}{\partial \tau} + \gamma_1 n_{G1} + k_1 C n_{G1} = 0 \tag{1}$$

$$\frac{\partial n_{M}}{\partial t} + \frac{\partial n_{M}}{\partial \tau} + \gamma_2 n_{M} + k_2 C n_{M} = 0 \tag{2}$$

$$\frac{\partial n_{MA}}{\partial t} + \frac{\partial n_{MA}}{\partial \tau} + \gamma_3 n_{MA} + \gamma_4 n_{MA} + \gamma_5 n_{MA} = 0 \tag{3}$$

Here

- $n_i(t,\tau)$ represents age density function, where t is the time and τ is the time spent in i^{th} phase
- $N_i(t) = \int_0^\infty n_i(t,\tau)\,\partial\tau$,where $N_i(t)$ is the number of cells in i^{th} phase at a given time

Population balance

Boundary conditions:

- $n_{G1}(t,\infty) = 0, n_M(t,\infty) = 0, n_{MA}(t,\infty) = 0$
- $n_{G1}(t,0) = 2\gamma_2 \int_0^\infty n_M(t,\tau)\partial\tau + \gamma_4 \int_0^\infty n_{MA}(t,\tau)\,\partial\tau$
  $= 2\gamma_2 N_m(t) + \gamma_4 N_{MA}(t)$
- $n_M(t,0) = \gamma_1 \int_0^\infty n_{G1}(t,\tau)\,\partial\tau = \gamma_1 N_{G1}(t)$
- $n_{MA}(t,0) = k_2 C \int_0^\infty n_M(t,\tau)\,\partial\tau = k_2 C N_M(t)$

Solution

$$\frac{d}{dt}(\mathsf{N}(t)) = \mathsf{A}\mathsf{N} \tag{4}$$

where

$$\mathsf{N} = \begin{bmatrix} N_{G1} \\ N_M \\ N_{MA} \end{bmatrix} \qquad \mathsf{A} = \begin{bmatrix} -(\gamma_1 + k_1 C) & 2\gamma_2 & \gamma_4 \\ \gamma_1 & -(\gamma_2 + k_2 C) & 0 \\ 0 & k_2 C & -(\gamma_3 + \gamma_4 + \gamma_5) \end{bmatrix}$$

This is an eigenvalue problem with solution

$$\hat{\mathsf{N}} = \sum_{i=1}^{n} c_i e^{\lambda_i t} \hat{\mathsf{z}}_{\mathsf{i}} \tag{5}$$

where

- n is the number of eigenvalues of A
- c_i is a constant
- λ_i is the i^{th} eigenvalue of A
- $\hat{\mathsf{z}}_{\mathsf{i}}$ is the eigenvector corresponding to λ_i

Solution

To find the eigenvalues we solve $|\mathbf{A} - \lambda \mathbf{I}| = 0$, a 3 deree polynomial equation in λ. If only one root is real then the solution is

$$\lambda_1 = k; \;\; \lambda_2 = a + ib; \;\; \lambda_3 = a - ib$$

$$\text{where } k, a, b \,\epsilon\, \mathbb{R}, \;\; i = \sqrt{-1}$$

The corresponding eigenvectors are

$$\hat{\mathbf{z}}_1 = \begin{bmatrix} u \\ v \\ w \end{bmatrix} \;\; \hat{\mathbf{z}}_2 = \begin{bmatrix} x_1 + iy_1 \\ x_2 + iy_2 \\ x_3 + iy_3 \end{bmatrix} \text{ and } \hat{\mathbf{z}}_3 = \begin{bmatrix} x_1 - iy_1 \\ x_2 - iy_2 \\ x_3 - iy_3 \end{bmatrix} = \bar{\hat{\mathbf{z}}}_2$$

Let

$$p = real(c_2) = real(c_3)$$

$$q = imag(c_2) = -imag(c_3)$$

Solving the Continuity Equations

Solving we get

$$N_{G1}(t) = c_1\, e^{kt}\, u + \\ 2e^{at}\,(\cos(bt)\, px_1 - \cos(bt)\, qy_1 - \sin(bt)\, py_1 - \sin(bt)\, qx_1) \tag{6}$$

$$N_{M}(t) = c_1\, e^{kt}\, v + \\ 2e^{at}\,(\cos(bt)\, px_2 - \cos(bt)\, qy_2 - \sin(bt)\, py_2 - \sin(bt)\, qx_2) \tag{7}$$

$$N_{MA}(t) = c_1\, e^{kt}\, w\,, + \\ 2e^{at}\,(\cos(bt)\, px_3 - \cos(bt)\, qy_3 - \sin(bt)\, py_3 - \sin(bt)\, qx_3) \tag{8}$$

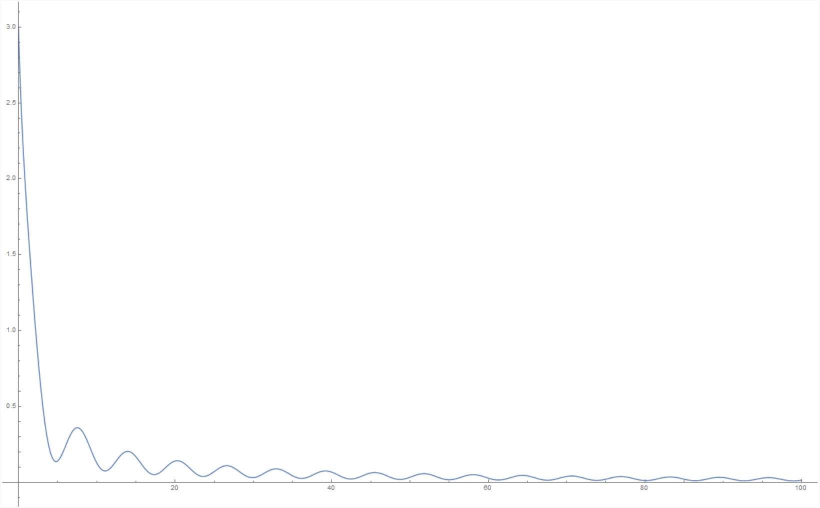


Figure 13: The ideal solution

Solving the Continuity Equations

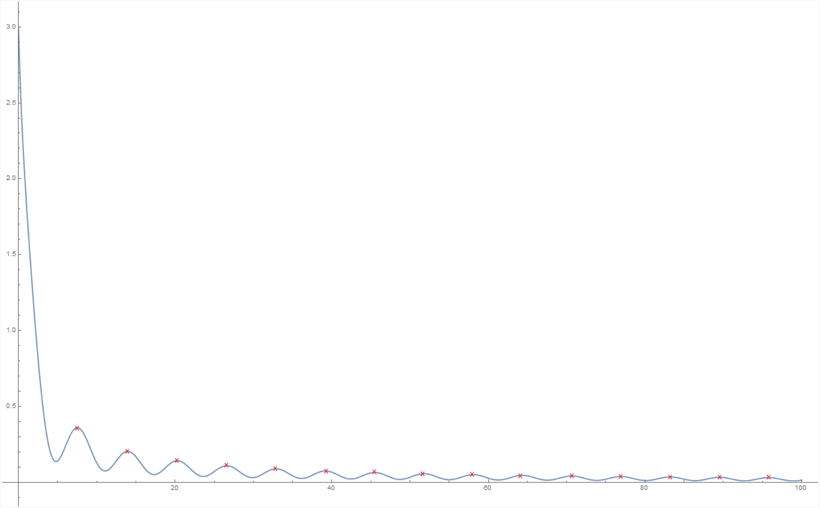


Figure 14: The ideal solution

Solution from MATLAB

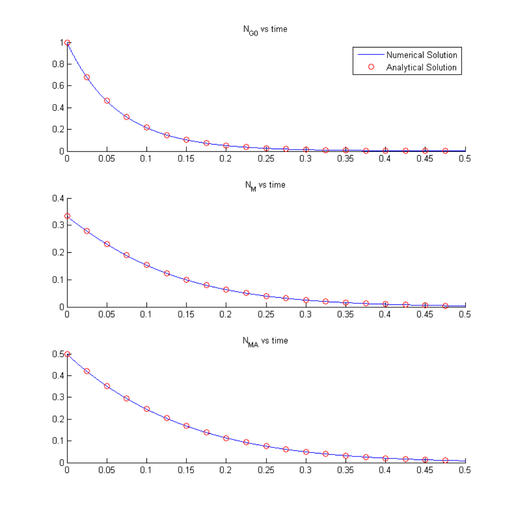


Figure 15: Solution

Solution from MATLAB

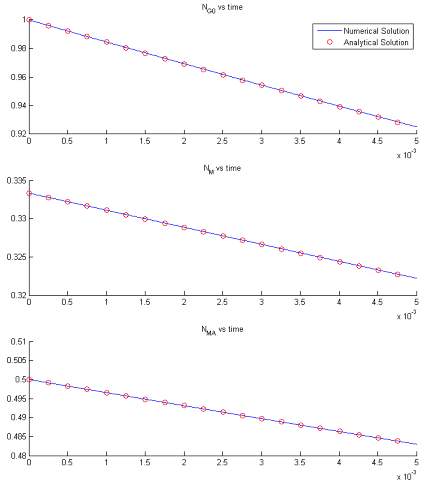


Figure 16: Solution with $C(t)$

Solution from MATLAB

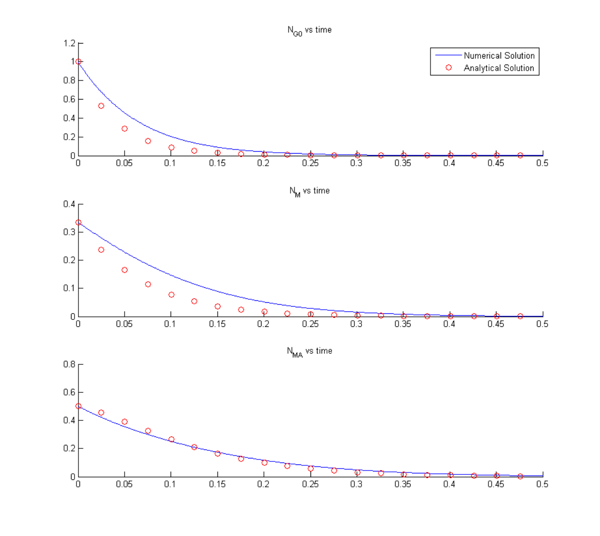
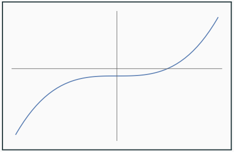
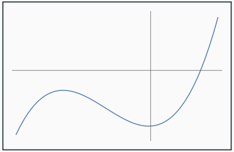


Figure 17: Solution with $C(t)$

Challenges

- We want $|\mathbf{A} - \lambda \mathbf{I}|$ to show the following behavior:





- Moreover for eqn (9) to show oscillations the following conditions must be met

$$\text{(i) } 2\,(a+b)\,\sqrt{(\,px_1 - qy_1)^2 + (\,py_1 + qx_1)^2} > k\,c_1\,u$$

$$\text{(ii) } a < 0$$

  - Many parameters
  - Parameters obtained from eigenvalues and eigenvectors
  - Difficult to control

Challenges

- The analytical solution fails on introducing time dependence
- Guessing values might result in a model which does not make biological sense

Conclusions

Summary

- With the help of our model one can predict a patient's susceptibility to VIPN, to decide proper dosage regimen
- The algorithms implemented can be applied for response prediction of many diseases
- Based on our model oscillations should be observed in the number of cells in M phase and G1 phase in cycle cell, and this could be used to time the doses for maximum efficacy
- However experimental data is required to verify our pharmacodynamic model for anti-mitotic drugs

Thank You

Bibliography

Backup Slides

Pharmacodynamic Model for anti-mitotic drug - Assumptions

- If the attack happens in the G1 phase the cell dies immediately, if it attacks in any other phase the cycle continues till M phase
- If the cell remains unaffected till M phase it will divide into two cells and either continue once again or exit the cycle.
- If the drug attacks during Mitosis or the drug attack leads to M phase, the cell enters Mitotic Arrest
- If the cell enters Mitotic Arrest, it may divide unequally or it may exit without division
- If the cell undergoes unequal division it cannot continue in cycle
- If the cell exits without dividing, it may once again start a cell cycle or enter Interphase Arrest or might die in Interphase

Drug Concentration model

For the j^{th} dose of i^{th} drug

$$C_i(t) = \begin{cases} \frac{y_{ij}}{\lambda_i}(1 - e^{-\lambda_i(t - t_{AD,ij})}) + C_{i,residual}(t), & \text{for } j^{th} \text{ application} \\ C_{i,residual}(t) = \sum_{j=1}^{PA} \frac{y_{ij}}{\lambda_i} e^{-\lambda_i(t - t_{AD,ij})}(e^{\lambda_i h_{ij}} - 1), & \text{b/w applications} \end{cases}$$

where

- $C_i(t)$ is the concentration of i^{th} drug at time t
- y_{ij} is the dose administration rate for the j^{th} dose of i^{th} drug
- λ_i is the decay constant for i^{th} drug
- h_{ij} is the dosage duration for the j^{th} dose of i^{th} drug
- $t_{AD,ij}$ time when the j^{th} dose of i^{th} drug started

Solving the Continuity Equations

Integrating eqn (1), (2) and (3) with respect to $\partial\tau$ and applying the boundary conditions we get

$$\frac{dN_{G1}}{dt} = 2\gamma_2 N_M + \gamma_4 N_{MA} - (\gamma_1 + k_1 C) N_{G1} \tag{9}$$

$$\frac{dN_M}{dt} = \gamma_1 N_{G1} - (\gamma_2 + k_2 C) N_M \tag{10}$$

$$\frac{dN_{MA}}{dt} = k_2 C N_M - (\gamma_3 + \gamma_4 + \gamma_5) N_{MA} \tag{11}$$

Challenges

- The function $|\mathbf{A} - \lambda \mathbf{I}|$ when expanded is

$$
\begin{aligned}
&-\lambda^3 + \lambda^2(-\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4 - Ck_1 - Ck_2) + \\
&\lambda(3\gamma_1\gamma_2 - \gamma_1\gamma_3 + \gamma_2\gamma_3 - \gamma_1\gamma_4 + \gamma_2\gamma_4 - C^2k_1k_2 - \gamma_1Ck_2 + \gamma_2Ck_1 - \gamma_3Ck_1 \\
&\quad - \gamma_3Ck_2 - \gamma_4Ck_1 - \gamma_4Ck_2) - \gamma_3C^2k_1k_2 - \gamma_4C^2k_1k_2 \\
&\quad - \gamma_1\gamma_3Ck_2 + \gamma_2\gamma_3Ck_1 + \gamma_2\gamma_4Ck_1 + 3\gamma_1\gamma_2\gamma_3 + 3\gamma_1\gamma_2\gamma_4
\end{aligned}
$$

- On differentiation with respect to λ

$$
\begin{aligned}
&-3\lambda^2 + 2\lambda(-\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4 - Ck_1 - Ck_2) + \\
&3\gamma_1\gamma_2 - \gamma_1\gamma_3 + \gamma_2\gamma_3 - \gamma_1\gamma_4 + \gamma_2\gamma_4 - C^2k_1k_2 - \gamma_1Ck_2 + \gamma_2Ck_1 - \gamma_3Ck_1 \\
&\quad - \gamma_3Ck_2 - \gamma_4Ck_1 - \gamma_4Ck_2
\end{aligned}
$$