Machine Learning Meets Systems Biology

Modeling Glucose-Insulin Response

Ariana Pahlavan, Emily Wang, Olivia Pfetsch, and William O'Brien

Introduction

Arianna

SEAS

Applied Math
Middle Eastern Studies Minor

Antitrust and M&A Law Part-time Employee at Pfizer

Emily

SEAS

Applied Math, Premedical Track

Olivia

SEAS

Applied Math, Earth and Environmental Engineering

Will

SEAS

Applied Math, CS and OR minors

Columbia T&F

ML Research in the Carleton Lab

ML Scientist at JPMorgan

OUTLINE

PART I: Background & Problem Context

PART II: The Minimal Model (Glucose Insulin Dynamic System)

PART III: Learning from Data with Machine Learning

PART IV: Comparison of Methods

PART V: Problem Impact

Primary References

Bergman, R N et al (1979). **Quantitative estimation of insulin sensitivity.** *The American journal of physiology* vol. 236,6 p. 667-77. doi:10.1152/ajpendo.1979.236.6.E667.

Boston, Raymond, et al (2003). **MINMOD Millennium: A Computer Program to Calculate Glucose Effectiveness and Insulin Sensitivity From the Frequently Sampled Intravenous Glucose Tolerance Test.** *Diabetes technology & therapeutics*, vol. 5, p. 1003-15. doi: 10.1089/152091503322641060.

Gilpin, William, et al (2020). **Learning dynamics from large biological data sets: Machine learning meets systems biology.** *Current Opinion in Systems Biology*, Vol. 22, p. 1-7, ISSN 2452-3100, https://doi.org/10.1016/j.coisb.2020.07.009.

Pacini, Giovanni, et al (1998). **Insulin sensitivity and glucose effectiveness: Minimal model analysis of regular and insulin-modified FSIGT.** *The American Journal of Physiology*. 274. E592-9. doi: 10.1152/ajpendo.1998.274.4.E592.

I

BACKGROUND

Why this problem is important to us?

Machine learning has met and improved many areas of society like financial services, online customer support, fraud detection, security and surveillance, automobiles but we believe that studying its interplay with systems biology will have a positive impact on mankind– save lives, prevent diseases, and ameliorate medical care overall

Our group has diverse interests from medical school to studying law to earth/environmental engineering to machine learning but what we all have in common is excitement for the future of modeling systems biology at a higher accuracy lower cost for the benefit of society

How is ML changing the way we model health?

How do dynamic systems fit into the new ML paradigm?

Background – ML Meets Dynamic Systems

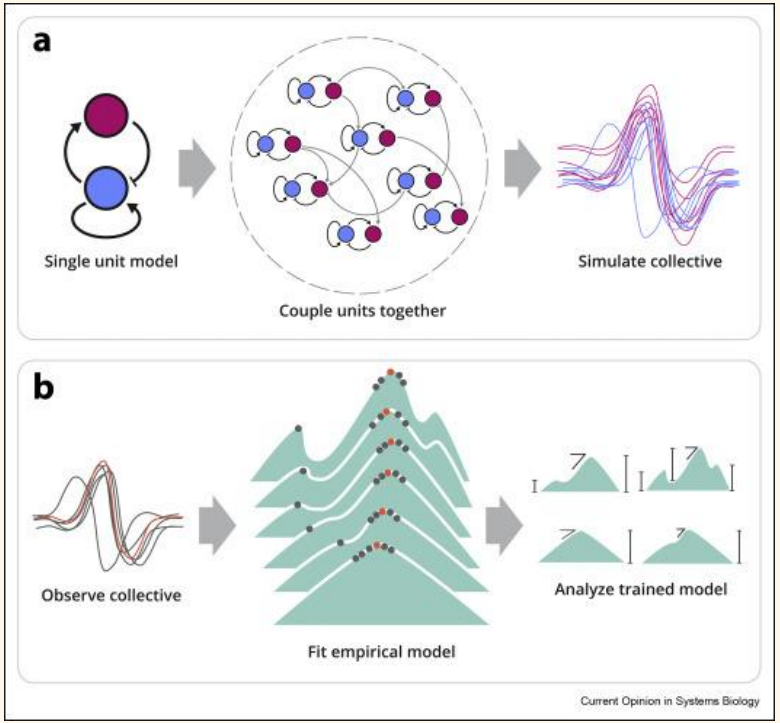


Fig. 1: Gilpin, W. et al (2020)

Bottom-up vs Top-down modeling

(a) Bottom-up: Experimental data are used to analyze collective dynamics of individual units in system

(b) Top-down: Measurements are fit to an empirical model (ML approach)

Background – ML Meets Dynamic Systems

Reductions in dimensions of complex systems in a bottom-up approach loses information compared to top-down approach

ML can lose interpretability

Blending the approaches may offer the best of both worlds

- Top-down to learn variables, then dynamic system built from learned relationships
- Symbolic regressions, fitting differential equations directly to observed data

Background – ML Meets Dynamic Systems

Unsupervised learning allows dynamics to be extracted from large biological data sets

Extracted dynamics can inform the development of mechanistic mathematical models

Machine learning models can be analyzed using tools from dynamical systems theory

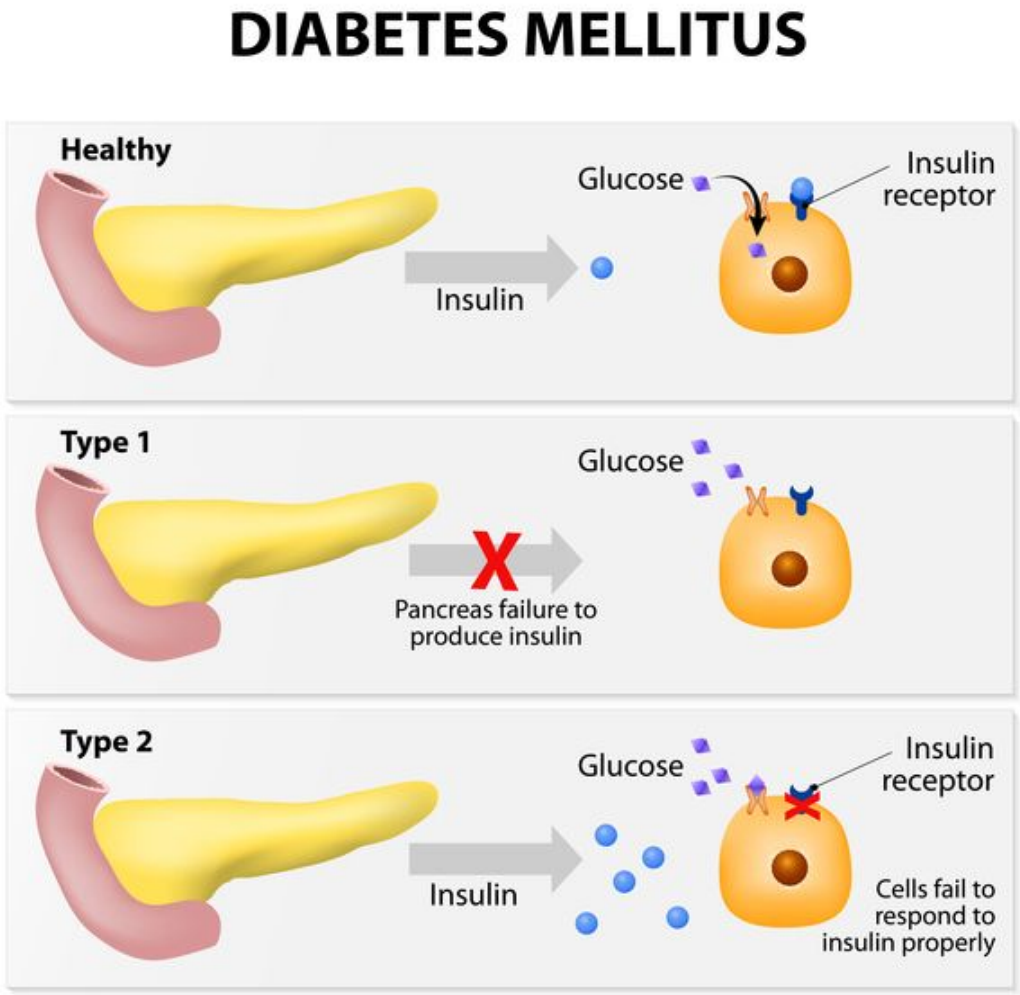
Gilpin, W. et al (2020)

Problem Context

Glucose is a form of sugar that circulates in our blood

Concentration of blood glucose is managed by the pancreas by producing Insulin

Hyperglycemia occurs when the pancreas does not produce enough insulin (Type I diabetes) or if insulin receptors become insensitive (Type II diabetes)



https://medlineplus.gov/genetics/condition/type-2-diabetes/

Problem Context

Frequently sampled intravenous glucose tolerance test (FSIGTT)

Glucose is injected into a fasting subject

Blood samples are collected at intervals of 2–10 minutes for 3 hours

Using FSIGTT data from past studies, we can quantify and model a subject's insulin responses to glucose

Problem Context

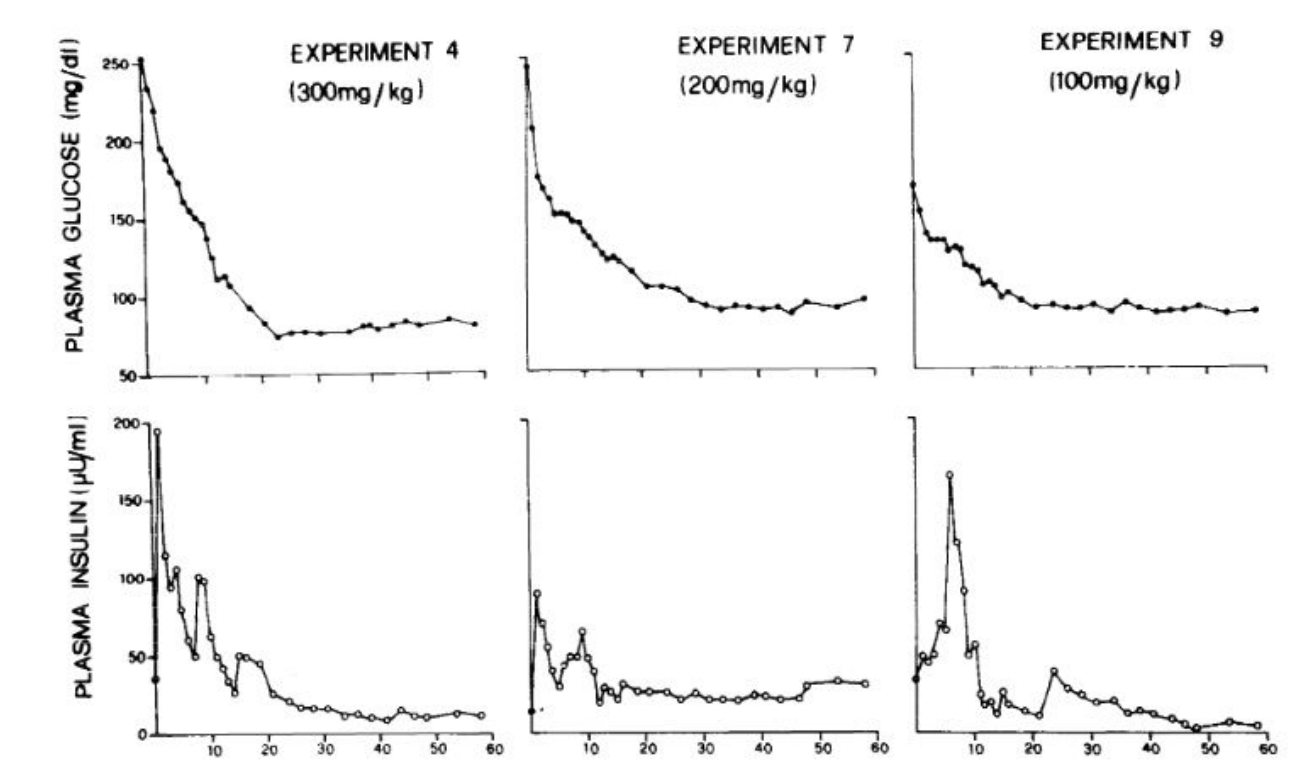


Fig. 2: Bergman R, N. et al (1979)

Problem Context

The organs that regulate blood-glucose concentration operate in a non-linear manner

Mathematical modeling can get valuable information on blood-glucose interactions

We apply bottom-up and top-down techniques to analyze the system

Problem Context

Graphic representation of the closed loop relationship between glucose, insulin secretion, and insulin action

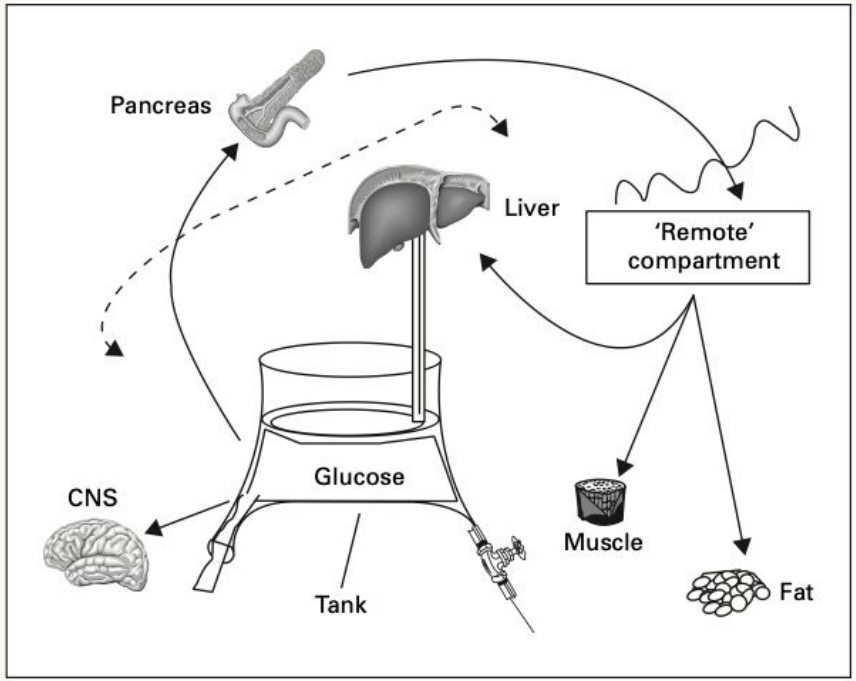


Fig. 3: Bergman R, N (2005)

Problem Context – Data

Data comes from Pacini, G. et al (1998)

The data we model contains glucose and insulin concentrations over 182 min for a single subject with normal insulin production and sensitivity

II

THE MINIMAL MODEL

THE MINIMAL MODEL

Bergman, R. et al apply Occam's Razor principle by seeking "the simplest model based upon known physiology that could account for the insulin–glucose relationship revealed in the data" (Bergman 2005, 9)

Derivation of the Minimal Model

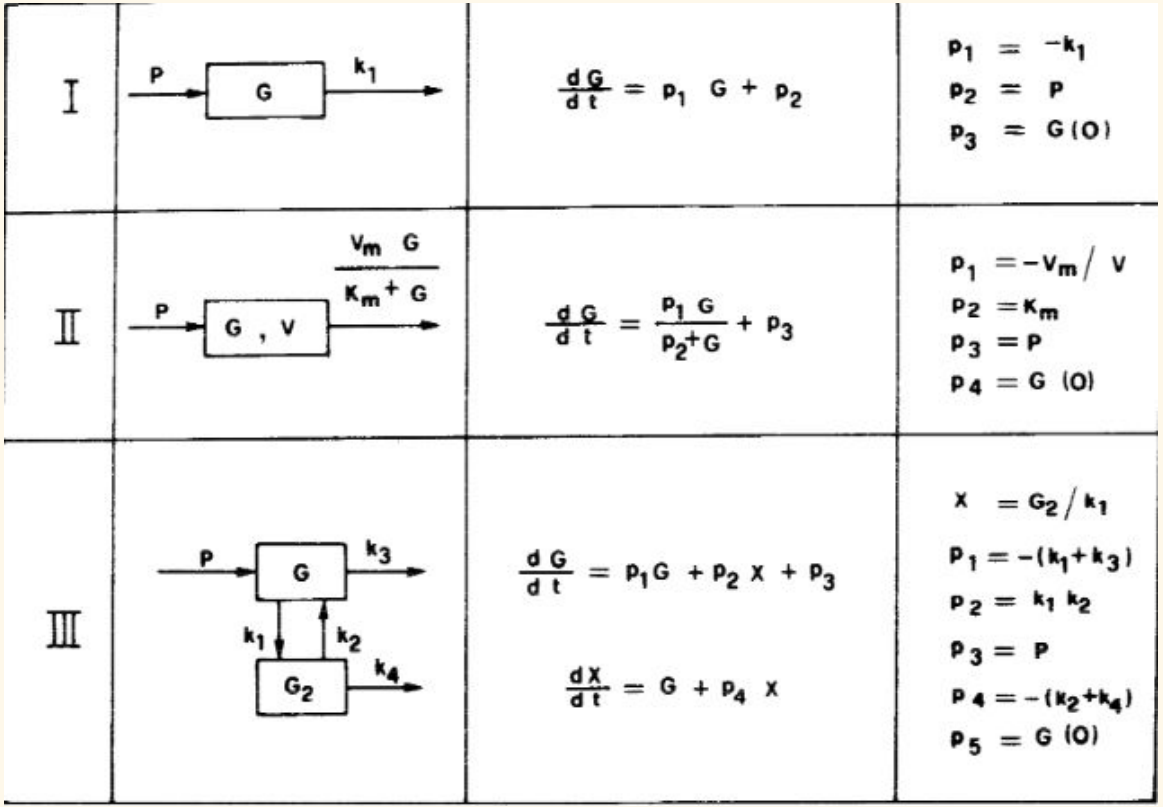


Fig. 4: Bergman R, N. et al (1979)

Derivation of the Minimal Model

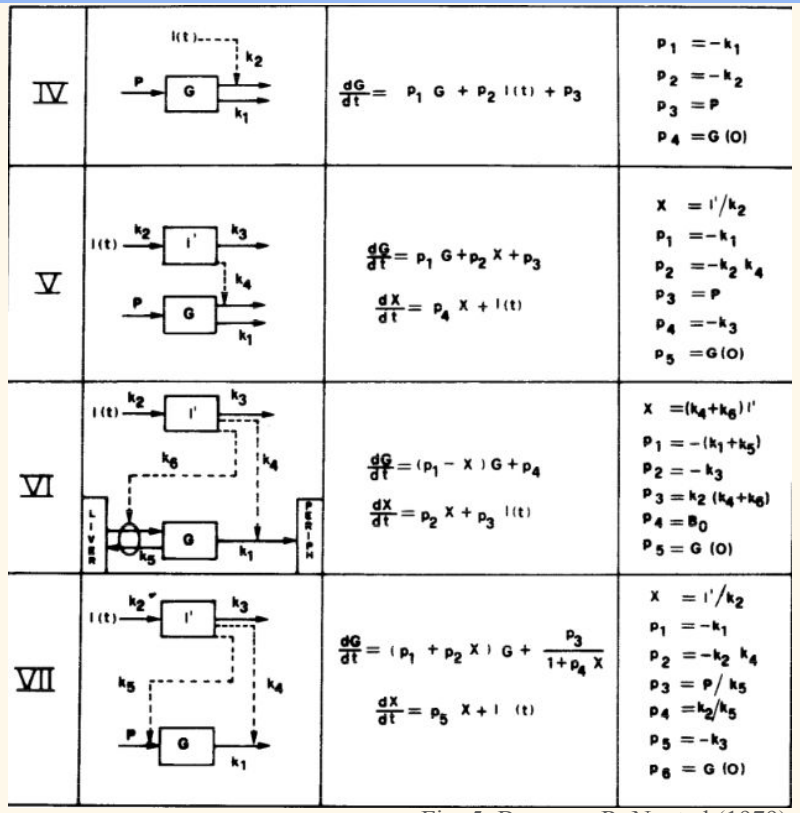


Fig. 5: Bergman R, N. et al (1979)

THE MINIMAL MODEL

glucose restoration rate

$$\begin{cases} \frac{dG}{dt} = -p_1[G(t) - G_b] - X(t)G(t) \\ \frac{dX}{dt} = p_3[I(t) - I_b] - p_2X(t) \end{cases}$$

rate of increase in remote insulin

Minimal Model Parameter Definitions

G - blood glucose concentration as a function of time

X - concentration of insulin in the tissue fluid as a function of time

I - concentration of insulin in the blood measured with input

G_b - equilibrium (basal) concentration of blood glucose measured before the trial

I_b - equilibrium (basal) concentration of blood insulin measured before the trial

p_i - parameters controlling rate of creation/removal of glucose and insulin

Breaking It Down

natural rate of glucose disappearance

interaction between glucose and insulin in tissue fluid

$$\begin{cases} \frac{dG}{dt} = -p_1[G(t) - G_b] - X(t)G(t) \\ \frac{dX}{dt} = p_3[I(t) - I_b] - p_2X(t) \end{cases}$$

rate of insulin diffusion between blood and tissue fluid

rate of insulin break down in tissue fluid

Interpreting Minimal Model Parameters

What can we learn from the model parameters (p_i)?

E - Glucose effectiveness

S - Insulin sensitivity

These quantities determine how effectively elevated glucose levels deplete glucose

Glucose Effectiveness

Tendency of elevated glucose levels to cause glucose depletion:

$$E = -\frac{\delta \dot{G}}{\delta G}$$

Taking this derivative leads to

$$E = p_1 + X$$

which arises from the Minimal Model

The glucose effectiveness index, S_G, arises when X approaches 0:

$$S_G = p_1$$

Insulin Sensitivity

How effectively does blood insulin change glucose effectiveness?

Measure of changes in glucose effectiveness w.r.t insulin concentration in the blood:

$$S = -\frac{\delta E}{\delta I}$$

Insulin Sensitivity Index

The insulin sensitivity index measures S when E and I are at a steady state:

$$\dot{G} = \dot{X} = 0 \quad \text{implies} \quad S_i = \frac{\delta E_{SS}}{\delta I_{SS}}$$

$$\frac{dX}{dt} = p_3[I(t) - I_b] - p_2 X(t)$$

If we set dX/dt = 0 as above and solve for X, we get

$$X_{SS} = \frac{p_3}{p_2} I_{SS} \quad \text{and since} \quad E = p_1 + X \quad \text{from earlier,}$$

$$S_I = \frac{\delta E_{SS}}{\delta I_{SS}} = \frac{\delta X_{SS}}{\delta I_{SS}} \quad \text{after taking the derivative of X w.r.t I}$$

This gives us an estimate of the insulin sensitivity index, $S_I = p_3 / p_2$

Solving Numerically - Runge-Kutta-Dormand-Prince

Runge-Kutta-Dormand-Prince is an adaptive ODE solver in the family of Runge-Kutta

Estimate the average slopes of the solution at given time intervals, use to approximate value of the functions at each time step

Fourth and fifth order approximations to estimate the solution and its error

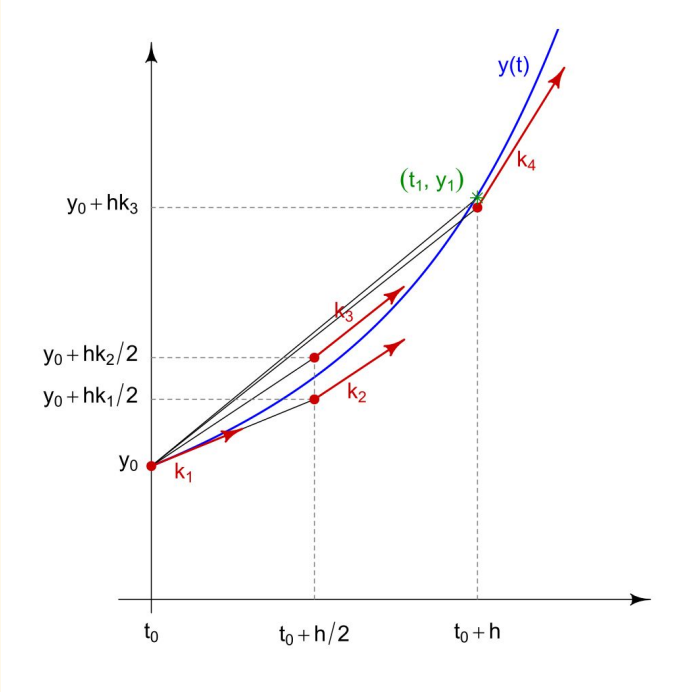


Fig. 6: HilberTraum (2017)

Solving Numerically - RKDP

Step 1:

$$k_1 = hf(t_k, y_k),$$
$$k_2 = hf\left(t_k + \frac{1}{5}h, y_k + \frac{1}{5}k_1\right),$$
$$k_3 = hf\left(t_k + \frac{3}{10}h, y_k + \frac{3}{40}k_1 + \frac{9}{40}k_2\right),$$
$$k_4 = hf\left(t_k + \frac{4}{5}h, y_k + \frac{44}{45}k_1 - \frac{56}{15}k_2 + \frac{32}{9}k_3\right),$$
$$k_5 = hf\left(t_k + \frac{8}{9}h, y_k + \frac{19372}{6561}k_1 - \frac{25360}{2187}k_2 + \frac{64448}{6561}k_3 - \frac{212}{729}k_4\right),$$
$$k_6 = hf\left(t_k + h, y_k + \frac{9017}{3168}k_1 - \frac{355}{33}k_2 - \frac{46732}{5247}k_3 + \frac{49}{176}k_4 - \frac{5103}{18656}k_5\right),$$
$$k_7 = hf\left(t_k + h, y_k + \frac{35}{384}k_1 + \frac{500}{1113}k_3 + \frac{125}{192}k_4 - \frac{2187}{6784}k_5 + \frac{11}{84}k_6\right).$$

$$y_{k+1} = y_k + \frac{35}{384}k_1 + \frac{500}{1113}k_3 + \frac{125}{192}k_4 - \frac{2187}{6784}k_5 + \frac{11}{84}k_6.$$

$$z_{k+1} = y_k + \frac{5179}{57600}k_1 + \frac{7571}{16695}k_3 + \frac{393}{640}k_4 - \frac{92097}{339200}k_5 + \frac{187}{2100}k_6 + \frac{1}{40}k_7$$

Error:

$$|z_{k+1} - y_{k+1}| = \left|\frac{71}{57600}k_1 - \frac{71}{16695}k_3 + \frac{71}{1920}k_4 - \frac{17253}{339200}k_5 + \frac{22}{525}k_6 - \frac{1}{40}k_7\right|$$

RKDP Pseudo Code

1. Define f’(t, y)
2. Declare array y of tmax doubles
3. While t < tmax
   a. Calculate k_1, k_2, k_3, k_4, k_5, k_6, k_7 as defined in the previous slide
   b. $y[t+1] \rightarrow y[t] + (35/384)\, k_1 + (500/1113)\, k_3 + (125/192)\, k_4 + (2187/6784)\, k_5 + (11/84)\, k_6$
   c. $z[t+1] \rightarrow y[t] + (5179/57600)\, k_1 + (7571/16695)\, k_3 + (393/640)\, k_4 + (-92097/339200)\, k_5 + (187/2100)\, k_6 + (1/40)\, k_7$
   d. $err \rightarrow | z_{t+1} - y_{t+1} |$
   e. $scale = 0.8 * | epsilon * y[t] / [err * (tmax - t)] |^{1/4}$
   f. $h \rightarrow scale * h$
4. Return y

Parameter Optimization - Least Squares

We use least squares optimization to find the parameter values p_1, p_2, p_3 that optimize the solution to the Minimal Model

Error function returns difference between the ODE solver output and training data, namely $\min\{||F(p)|| : p \epsilon R^n\}$ where $F \epsilon R^n$ and range in R^m for n=3 parameters and m data points:

$$F = \begin{pmatrix} f_1(p) \\ f_2(p) \\ \vdots \\ f_m(p) \end{pmatrix}$$

where f_i represents the difference between training data y and the solved ODE G at parameters p:

$$f_i : R^n \to R, f_i(p) = y_i - G(t_i, p)$$

More et al (1980)

Least Squares Optimization

Estimate initial parameters $p_0 = [\, p_{0,1}, p_{0,2}, p_{0,3} \,]$

Algorithm determines a correction k to p that decreases the residuals of F s.t. $||F(p+k)|| < ||F(x)||$

k is determined by

$$\min \|f + Jk\| : \|Dp\| \leq \Delta$$

where J is the Jacobian of F at p, and D is a diagonal scaling matrix

Until the criteria based on an error tolerance XTOL is met,

$$\Delta \leq \mathrm{XTOL} \cdot \|Dp\|$$

the step bound Δ is decreased and D and J are updated accordingly until we find optimal p

More et al (1980)

III

LEARNING FROM DATA WITH ML

Learning Models from the Data

For systems of ODEs and PDEs that cannot be solved analytically, there is a field of literature that explores the use of neural networks to approximate solutions

We first build a simple support vector regression, modeling our subject's insulin blood concentration increase in response to the glucose injection

Next we try using a neural network to approximate a solution function to the ODE system

Support Vector Regression

Given the low dimensionality of our data, we decided to use an SVR to capture the nonlinear relationship between glucose concentration and insulin using a polynomial kernel trick

$$
\begin{aligned}
\phi(a)^T \cdot \phi(b) &= \begin{pmatrix} a_1^2 \\ \sqrt{2}a_1a_2 \\ a_2^2 \end{pmatrix}^T \cdot \begin{pmatrix} b_1^2 \\ \sqrt{2}b_1b_2 \\ b_2^2 \end{pmatrix} \\
&= a_1^2b_1^2 + 2a_1b_1a_2b_2 + a_2^2b_2^2 \\
&= (a_1b_1 + a_2b_2)^2 \\
&= \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}^T \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \right)^2 = (a^T \cdot b)^2
\end{aligned}
$$

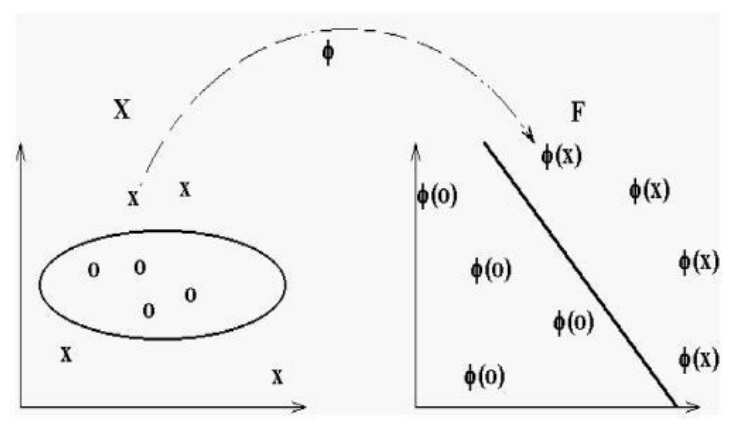


Fig. 7: Polynomial Kernel Trick where data is mapped to a linear space (Image Public Domain)

Support Vector Regression

Define the following variables:

α - vector of coefficients for SVR best fit

Q - a semidefinite matrix of our transformed data phi(x)

e - a vector of 1's

C - upper bound of all variables

ϵ - margin of tolerance

$$\min \tfrac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \epsilon \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)$$

$$+ \sum_{i=1}^{l} z_i(\alpha_i - \alpha_i^*) \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0$$

$$0 \le \alpha_i,\ \alpha_i^* \le C, \qquad i = 1, \ldots, l$$

where $Q_{ij} = \phi(x_i)^T \phi(x_j)$.

Lin, Chih-Jen (2001)

SVR Pseudo Code – Decomposition Method

Algorithm I.1—Decomposition Method:

1) Given a number $q \leq l$ as the size of the working set. Find α^1 as the initial solution. Set $k = 1$.
2) If α^k is an optimal solution of (1), stop. Otherwise, find a working set $B \subset \{1, \ldots, l\}$ whose size is q. Define $N \equiv \{1, \ldots, l\} \backslash B$ and α_B^k and α_N^k to be subvectors of α^k corresponding to B and N, respectively.
3) Solve the following subproblem with the variable α_B:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\alpha_B^T Q_{BB}\alpha_B - (e_B - Q_{BN}\alpha_N^k)^T\alpha_B \\ & 0 \leq (\alpha_B)_i \leq C, \qquad i = 1, \ldots, q, \\ & y_B^T\alpha_B = -y_N^T\alpha_N^k \end{aligned} \tag{2}$$

where $\begin{bmatrix} Q_{BB} & Q_{BN} \\ Q_{NB} & Q_{NN} \end{bmatrix}$ is a permutation of the matrix Q.

4) Set α_B^{k+1} to be the optimal solution of (2) and $\alpha_N^{k+1} \equiv \alpha_N^k$. Set $k \leftarrow k + 1$ and go to Step 2).

Lin, Chih-Jen (2001)

Solving Systems of ODEs with Deep Learning

Can we train a neural network to approximate solutions to our ODE system?

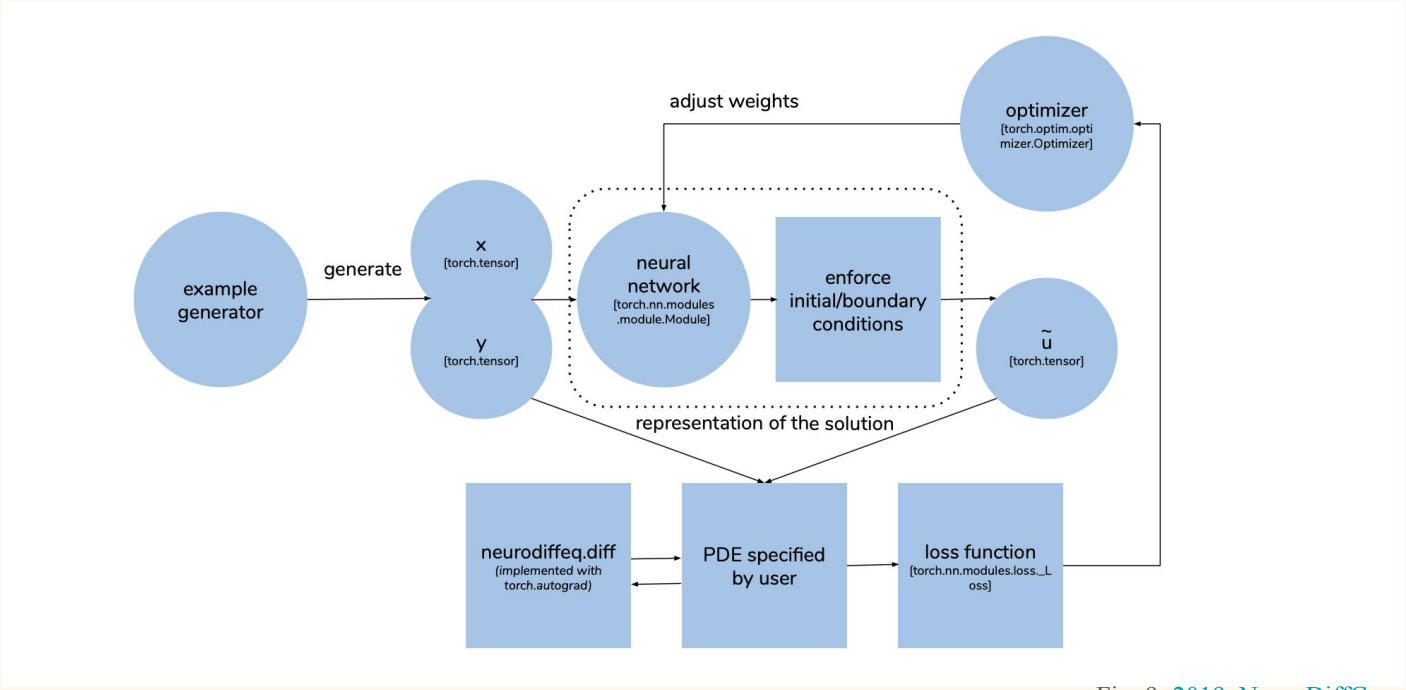
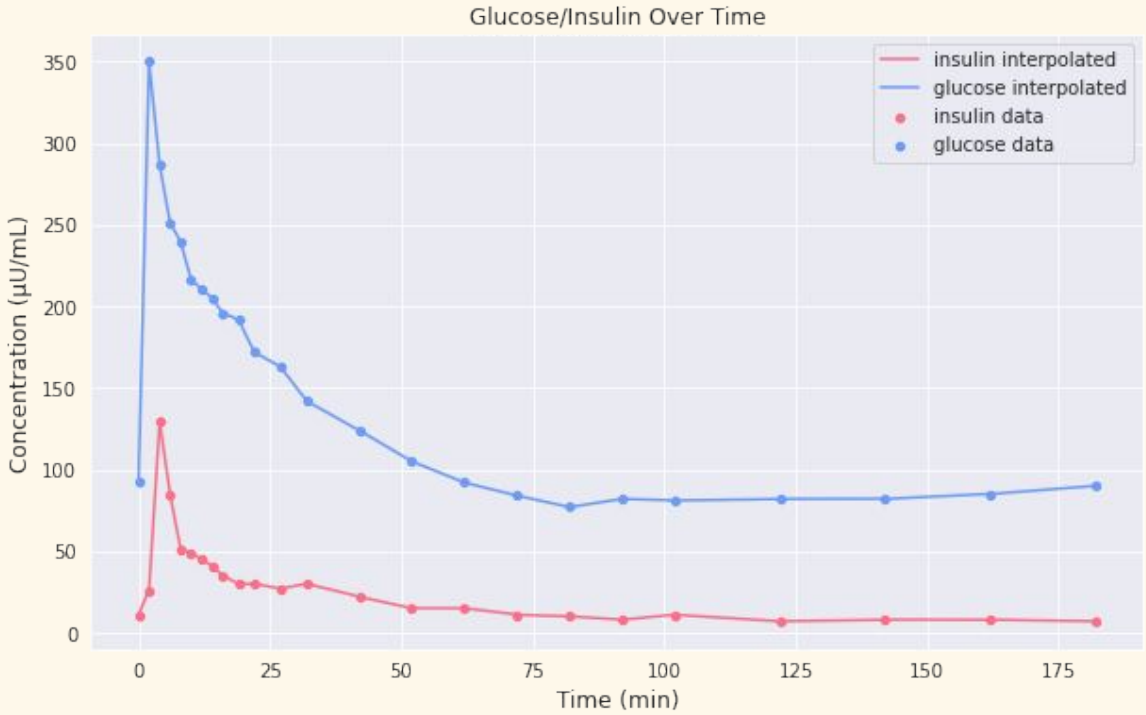


Fig. 9: 2019, NeuroDiffGym

IV

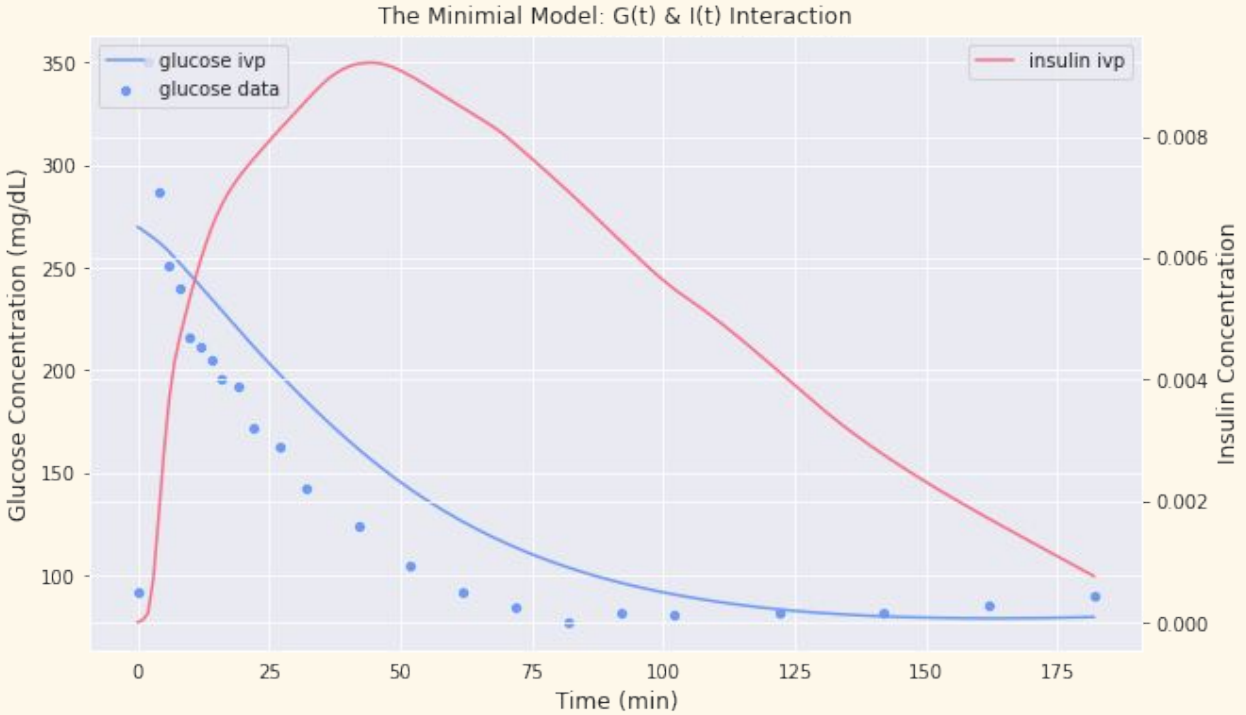
COMPARISON & RESULTS

Graphing Glucose-Insulin Interaction



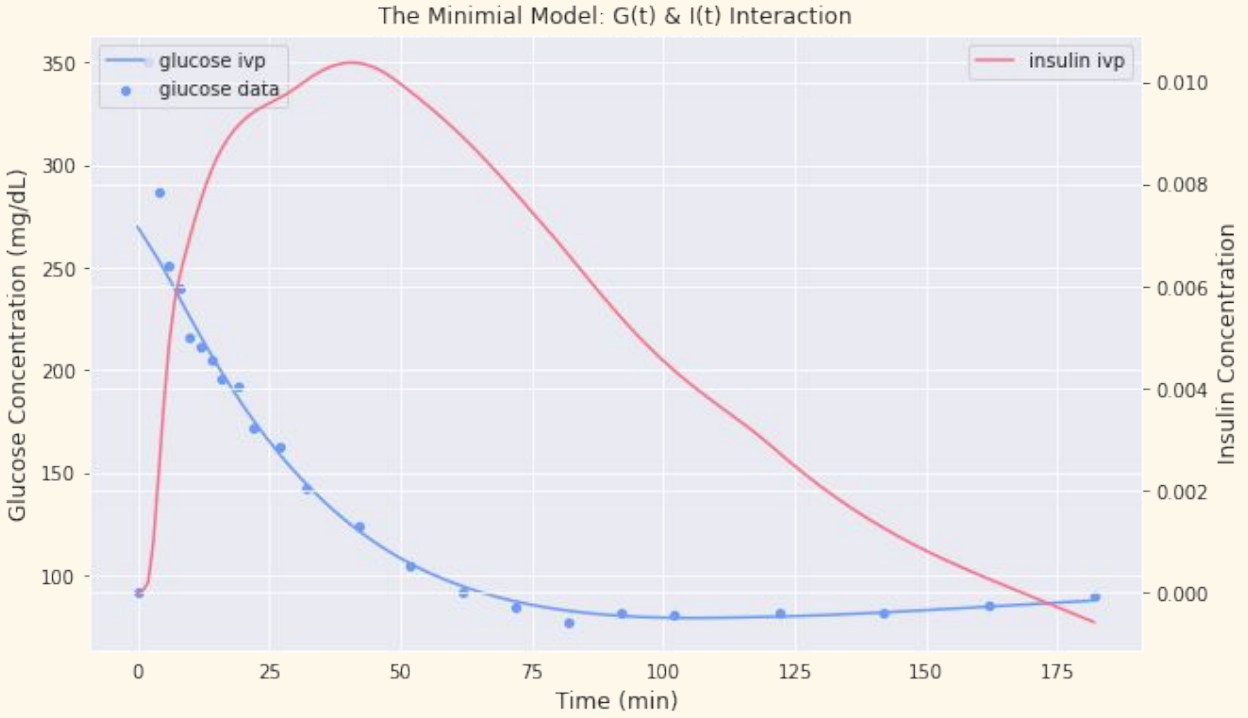
Data sourced from Pacini, Giovanni, et al (1998)

Solving Minimal Model, Guessing Parameters



Data sourced from Pacini, Giovanni, et al (1998)

Solving Minimal Model, Least Squares Parameter Optimization



Data sourced from Pacini, Giovanni, et al (1998)

Glucose Minimal Model Parameter Ranges

TABLE 1. GLUCOSE MINIMAL MODEL INDICES, THEIR UNITS, AND TYPICAL NORMAL VALUES AND NORMAL RANGES

Clinical index			*Typical normal*	
Name	*Abbreviation*	*Unit*	*Value*	*Range*
Glucose effectiveness	S_G	min^{-1}	2.2e-2	(1.2e-3, 4.5e-2)
Insulin sensitivity	S_I	$(mU/L)^{-1}\ min^{-1}$	2.0e-4	(5.0e-5, 2.2e-3)
	P2	min^{-1}	5.0e-2	(1.3e-3, 2.0e-1)
	P3	$(mU/L)\ min^{-2}$	2.1e-5	(5.4e-7, 8.0e-5)
Basal glucose	G_b	$mg\ dL^{-1}$	84	(65, 103)
Distributed glucose concentration at time 0	G_0	$mg\ dL^{-1}$	200	(150, 400)
Basal insulin	I_b	$mU\ L^{-1}$	10	(1,32)
Acute insulin response to glucose	AIR_g	$mU\ L^{-1}\ min^{-1}$	800	(45, 3700)
Disposition index	*DI*		8.0e-2	(1.6e-2, 1.0)
Glucose effectiveness at zero insulin	*GEZI*	min^{-1}	1.8e-2	(NA)
Insulin-attributable glucose disposal	*IAGD*	%	10	(0, 90)
β-Cell function	β-Cell function	mU/mM	170	(30, 1,440)
Insulin resistance		$mM\ mU\ L^{-2}$	2.0	(0.4, 8)
Insulin action	*X*	min^{-1}	0.01	(0, 0.03)
Apparent volume of glucose distribution	V_g	dL	140	(30, 292)

NA, not applicable.

Fig 8: Raymond et al (2003)

Solving Minimal Model, Least Squares Parameter Optimization

Raymond et al (2003) studied 7F and 9M patients (age ϵ [25, 64], 41±3; BMI ϵ [20.8, 29.4], 24.6 ± .07) found the following ranges of insulin and glucose sensitivity indices:

$S_I \ \epsilon$ [1.2e-3, 4.5e-2]

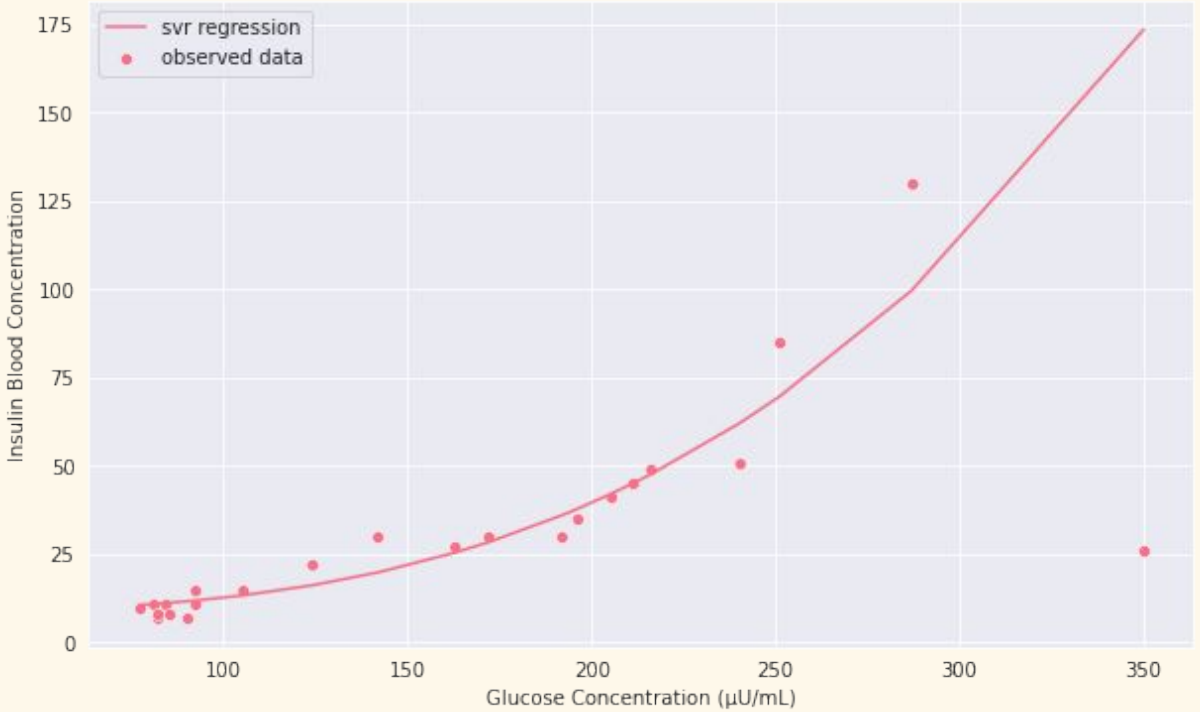
$S_G \ \epsilon$ [5.0e-5, 2.2e-3]

Our results from the Bergman et al (1979) data found both were in the normal range:

Insulin Sensitivity Index (S_I): $8.4e10^{-4}$ $min^{-1} \cdot \mu U^{-1} \cdot ml$

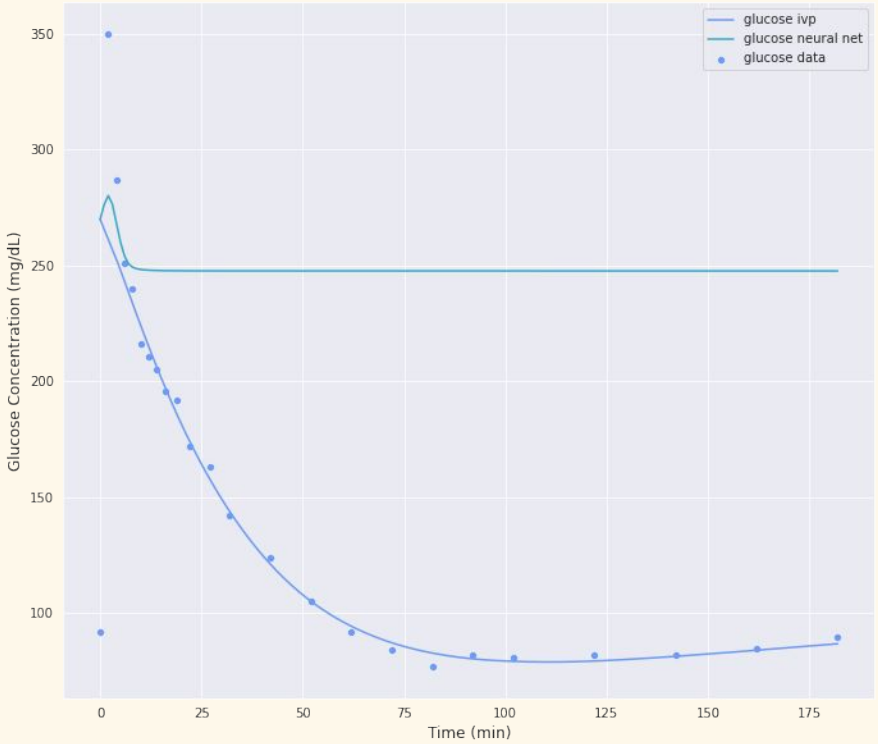
Glucose Sensitivity Index (S_G): 0.023 min^{-1}

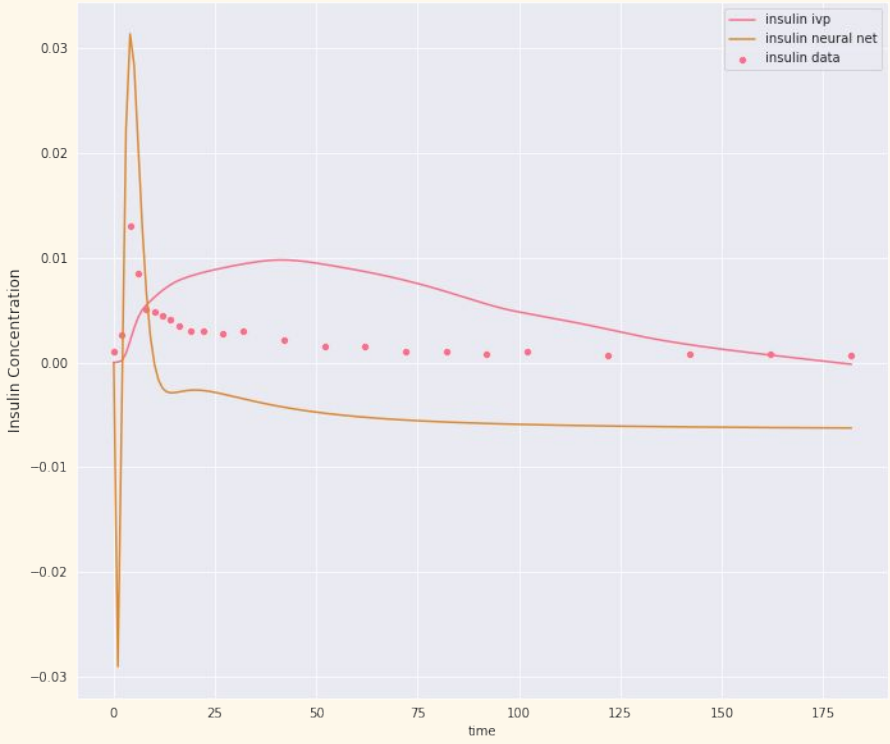
SVR Modeling Glucose Against Insulin Response



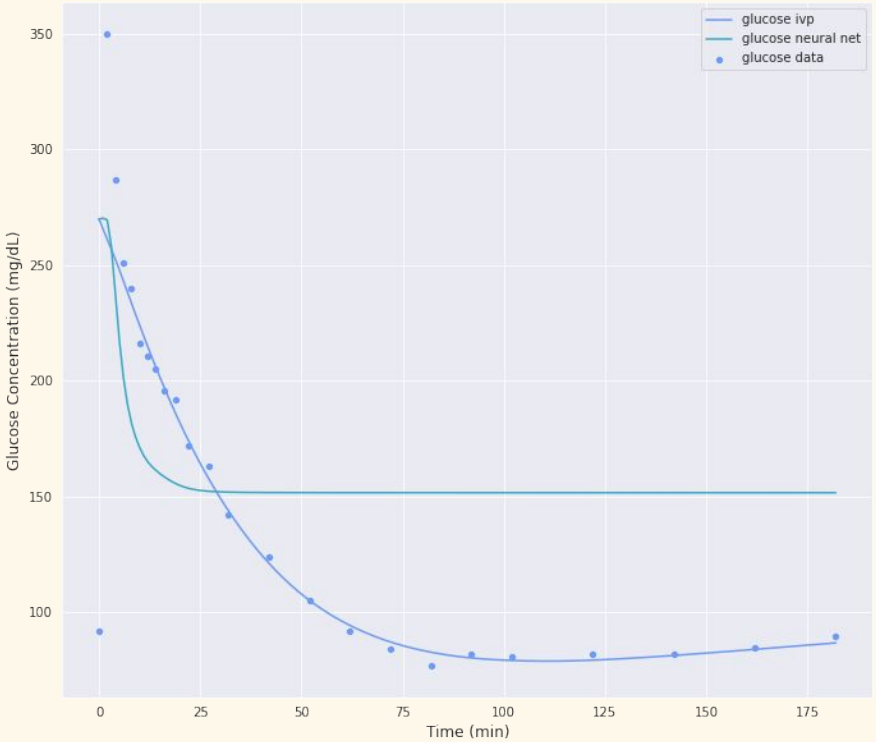
Data sourced from Pacini, Giovanni, et al (1998)

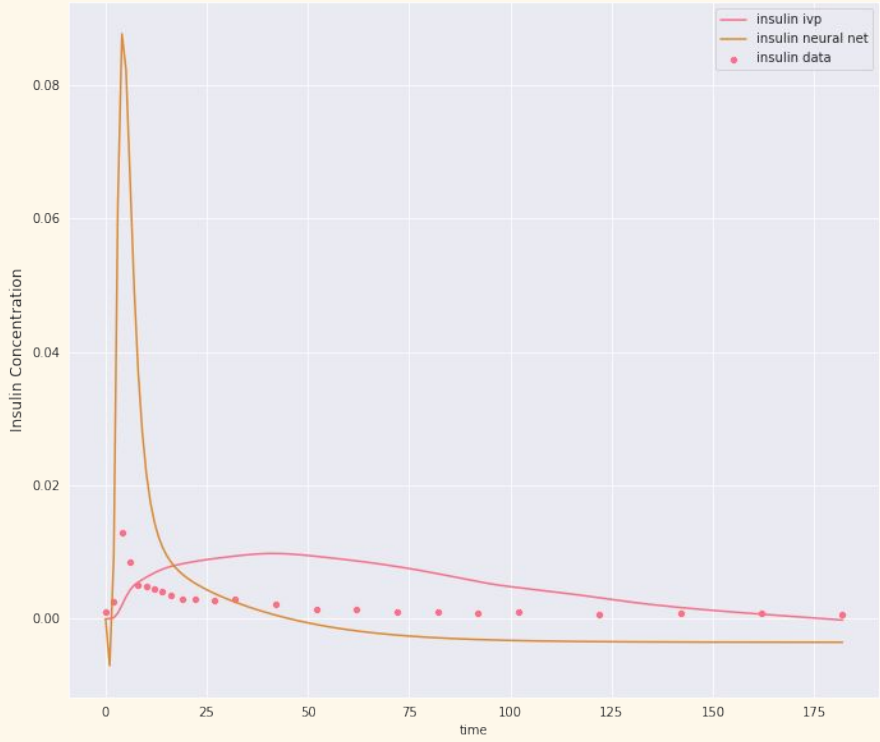
Neural Net ODE Solution – 3,000 Epochs of Training



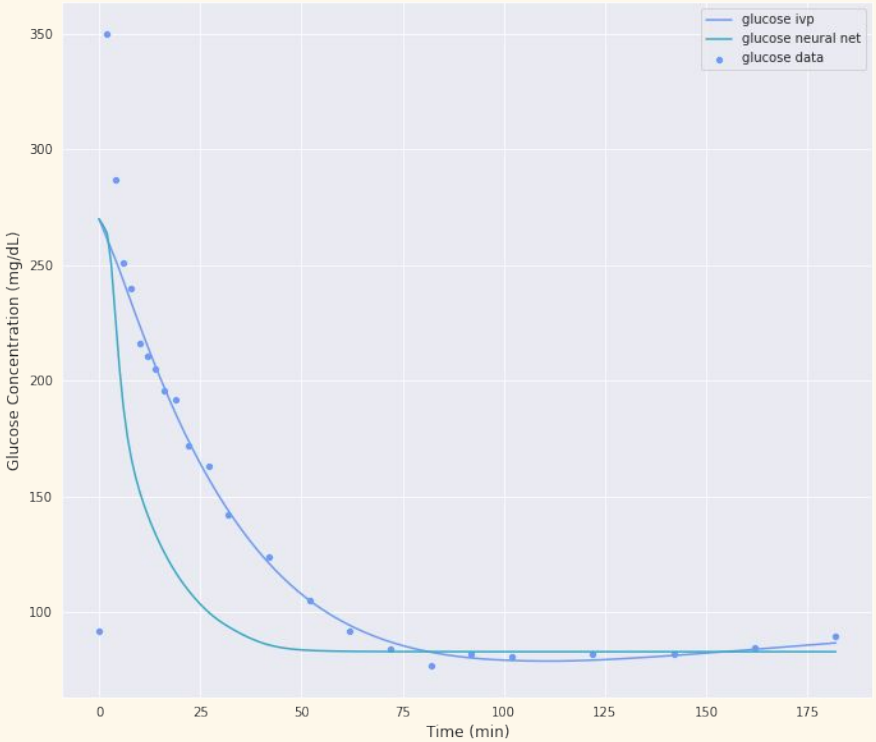


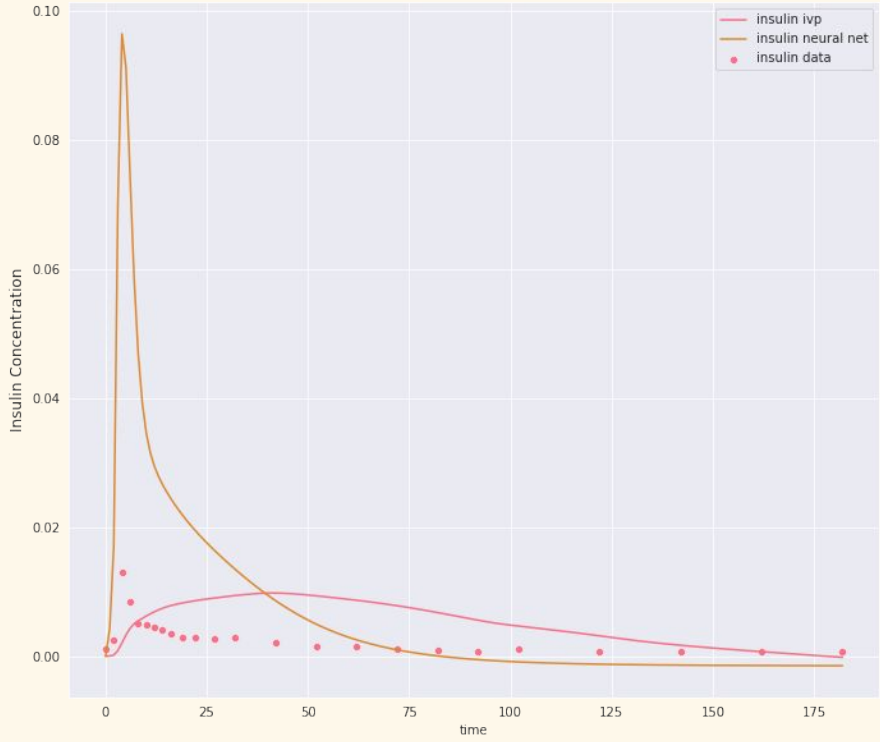
Neural Net ODE Solution – 10,000 Epochs of Training





Neural Net ODE Solution – 20,000 Epochs of Training





Colab Notebook

V

PROBLEM IMPACT

How is ML changing the way we model health?

How do dynamic systems fit into the new ML paradigm?

How is ML changing the way we model health?

Personalized healthcare (PHC): ML in the near future can improve medicine with increased abilities to analyze data. Each patient has bulks of data points: x-ray results, vaccinations, blood samples, vital signs, DNA sequence, current medication, family history, medical history, race, gender, menstrual cycle – all data points that have not been utilized to perfect medical treatment in the status quo yet

PHC is proactive as opposed to reactive

ML can resolve healthcare's "Missing Women" crises

8 major problems in modern healthcare

How do dynamic systems fit into the new ML paradigm?

Dynamic systems and ML are not mutually exclusive – there are overlaps in the modeling techniques that can have nice synergies

A neural network can fit high dimensional, complex DEs faster than some numerical solvers and can provide differentiable function solutions while still satisfying the system boundaries

Classic numerical solvers can be more stable, interpretable, and work well on most dynamic systems

Both are valuable tools in the modeling toolbox, but knowing when to use each is essential

Our models are driven and informed by data.

How are the biases in our healthcare systems causing societal harm?

How are the biases in our healthcare systems causing societal harm?

Women seem to have diabetes less often than men (although this is possibly inaccurate given how we established that they are often misdiagnosed/undiagnosed)

Females often have more serious complications and a greater risk of death

- Common complications for all diabetics: amputation, neuropathy, retinopathy, cardiovascular disease and kidney disease
- 45% of males with diabetes get erectile dysfunction
- **Women** have a much greater chance of heart disease, kidney disease and depression
- **Menopause (hormones)**: increase blood glucose more, increase weight and cause sleeping problems
    - These risks further exacerbate pre-existing health conditions

How are the biases in our healthcare systems causing societal harm?

Women are poorly diagnosed AND treated.

Research on disease in women is dwarfed by research in males - women often left out of medical studies, even female-prevalent diseases or treatments for females (if researchers even bother to make drugs for female issues)

Often misdiagnosed since symptoms appear differently in women (Ex. abdominal pain is often thought of as pregnancy, and chest pain often radiates to back pain or breathlessness etc.)

Women are more likely to adversely react to drugs - since drugs weren't made/tested with them in mind

Women are more likely to overdose - due to "gender-neutral" dosage standards

A cliche women run into when (often male) doctors are unable to diagnose them:

"You're crazy, [the pain] is *all in your head*."

How are the biases in our healthcare systems causing societal harm?

Invisible Women, Caroline Criado Perez:

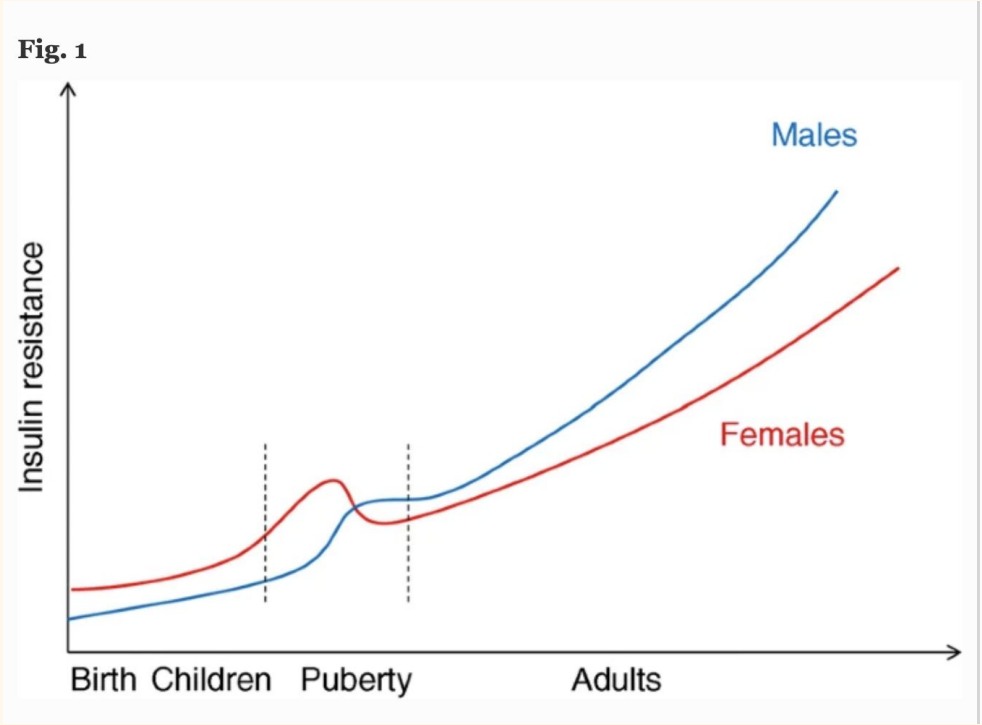
“Researchers have found sex differences in every tissue and organ system in the human body, as well as in the ‘prevalence, course and severity’ of the majority of common human diseases.”

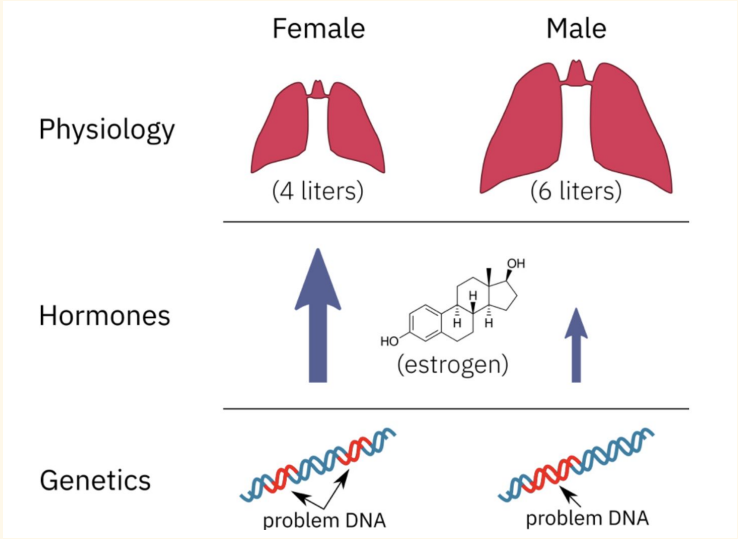
“A recent study also found a significant sex difference in the ‘expression of a gene found to be important for drug metabolism’.25”

“Because of their routine exclusion from clinical trials we lack solid data on how to treat pregnant women for pretty much anything.”

“[D]rugs [are not] tested in women at different stages in their menstrual cycles. … When women are included in trials at all, they tend to be tested … when hormone levels are at their lowest – i.e. when they are superficially most like men.”

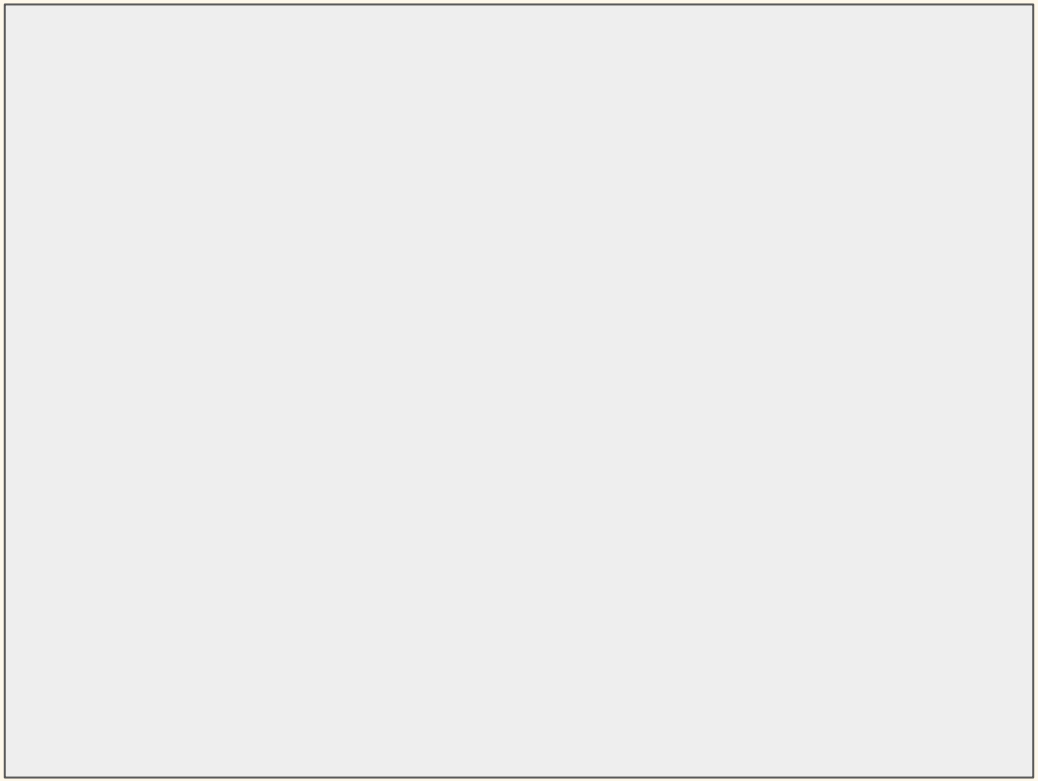
Sex Differences





Huebschmann (2019)

Quick message and concluding thought



Future Plans

Arianna

Law School

Return to Pfizer

Also interested in human rights work

Emily

Medical School or Graduate School (Biostatistics)

Interested in intersecting medicine and math

Olivia

Consulting for research institutes at Huron Consulting

Interested in women's rights work

Will

Returning to ML at JPMorgan

Interested in grad school after working a bit

Future Plans for this project: to continue to follow and read about new findings in this domain as we think in the future there will be more tech projects like Google's DeepMind investigating this space

Secondary References

Bergman R, N (2005). **Minimal Model: Perspective from 2005.** *Hormone Research,* 2005, Vol. 64 (suppl 3), p. 8-15, https://doi.org/10.1159/000089312.

Criado-Perez, Caroline. **"The Drugs Don't Work."** *Invisible Women: Data Bias in a World Designed for Men*, Abrams Press, New York, NY, 2021, pp. 285-315.

Dormand, J.R. and P.J. Prince (1980). **A family of embedded Runge-Kutta formulae.** *Journal of Computational and Applied Mathematics*, Vol. 6, Iss. 1, p. 19-26, ISSN 0377-0427, https://doi.org/10.1016/0771-050X(80)90013-3.

Lin, Chih-Jen (2001). **On the convergence of the decomposition method for support vector machines.** *IEEE Transactions on Neural Networks*, vol. 12, no. 6, p. 1288-1298, doi: 10.1109/72.963765.

More, Jorge J. et al. **User Guide For MINPACK-1.** Argonne National Laboratory, Argonne, IL, 1980, pp. 17-23. https://cds.cern.ch/record/126569/files/CM-P00068642.pdf