

Annual Review of Statistics and Its Application

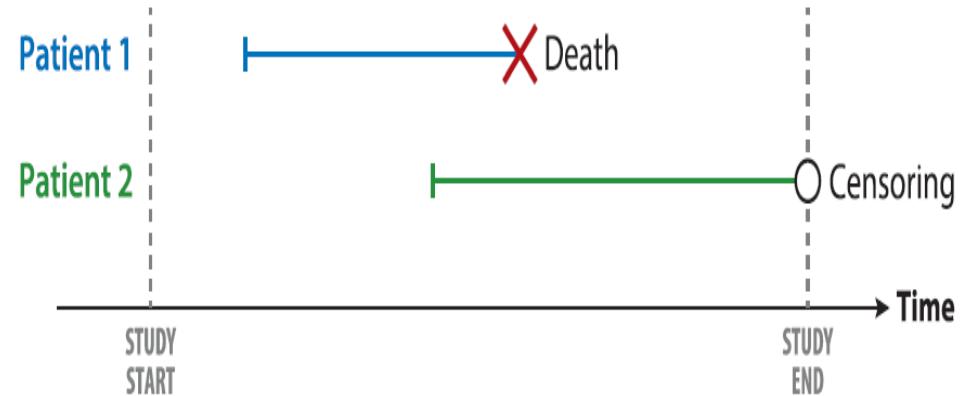
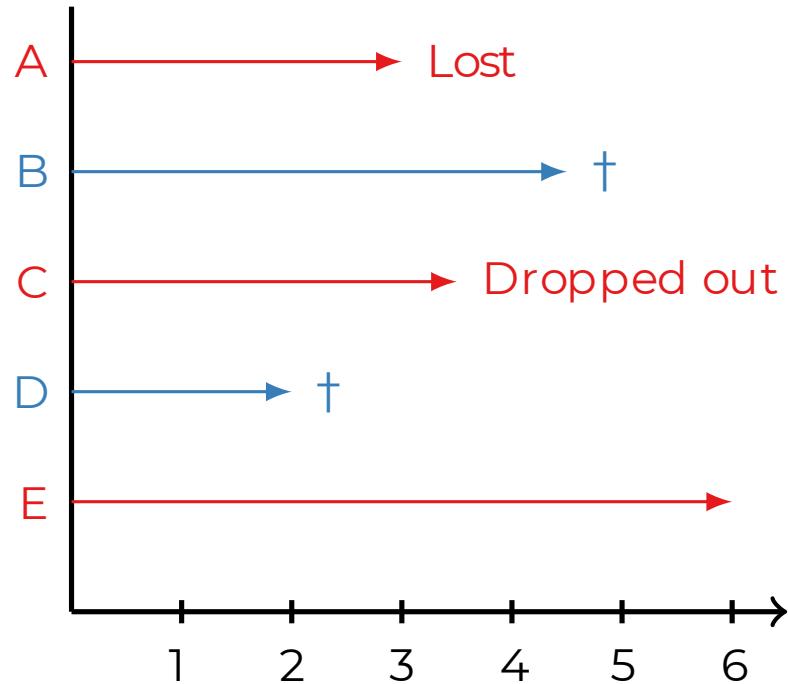
High-Dimensional Survival
Analysis: Methods and
Applications

Stephen Salerno and Yi Li

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA;
email: yili@umich.edu

Presented by : Siamak Salimi

Survival analysis



$$S(t) = P(T > t),$$

Survival analysis is a type of regression problem .

Hazard function

$h(t)$ Hazard function , denotes an approximate probability that an event occurs in the small time interval .

$[t; t + \Delta t[, \quad$: small time interval

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \geq 0.$$

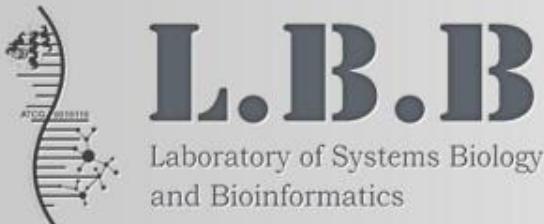
Conditional failure rate

Conditional mortality rate

Instantaneous failure rate

A KEY QUESTION ?

How to perform variable selection ?



High-dimensional survival models

J. R. Statist. Soc. B (1996)
58, No. 1, pp. 267–288

Regression Shrinkage and Selection via the Lasso

By ROBERT TIBSHIRANI†

University of Toronto, Canada

[Received January 1994. Revised January 1995]

› [J Stat Softw.](#) 2010;33(1):1-22. Journal of Statistical Software

Regularization Paths for Generalized Linear Models via Coordinate Descent

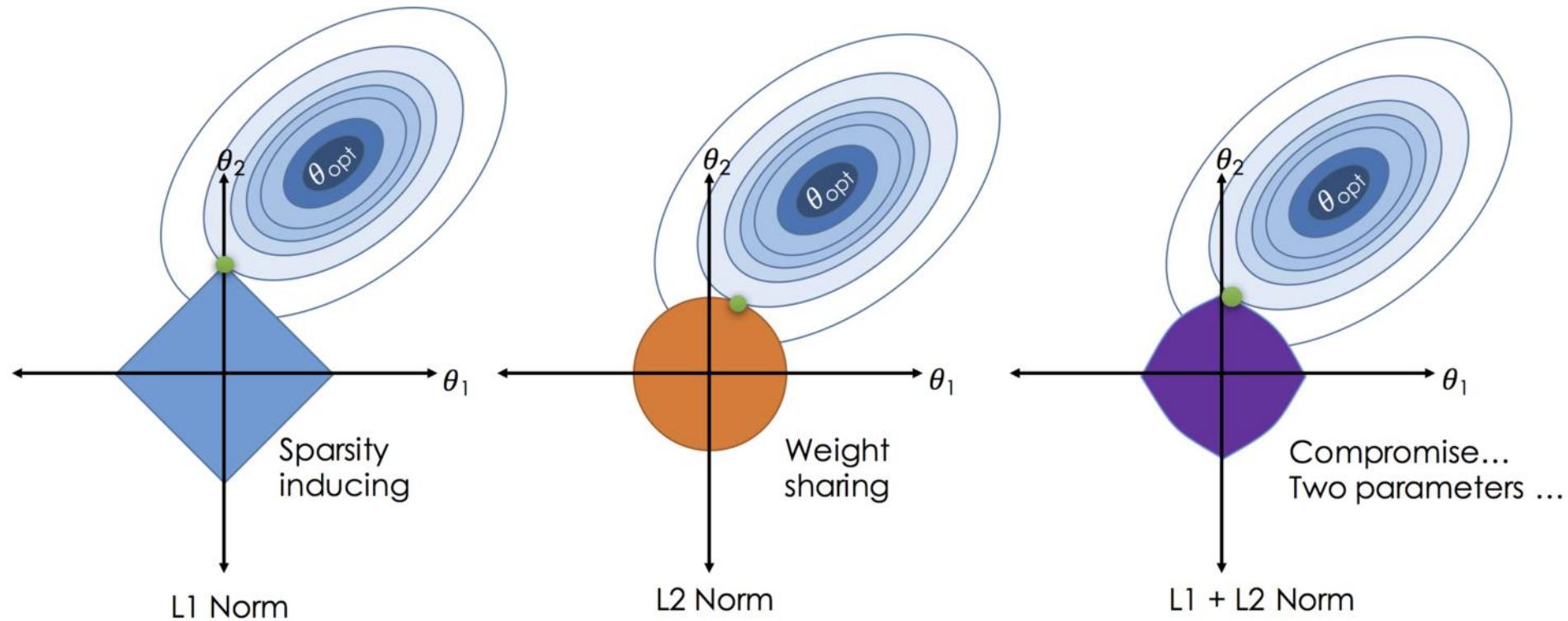
Jerome Friedman ¹, Trevor Hastie, Rob Tibshirani

Affiliations + expand

PMID: 20808728 PMCID: [PMC2929880](#)



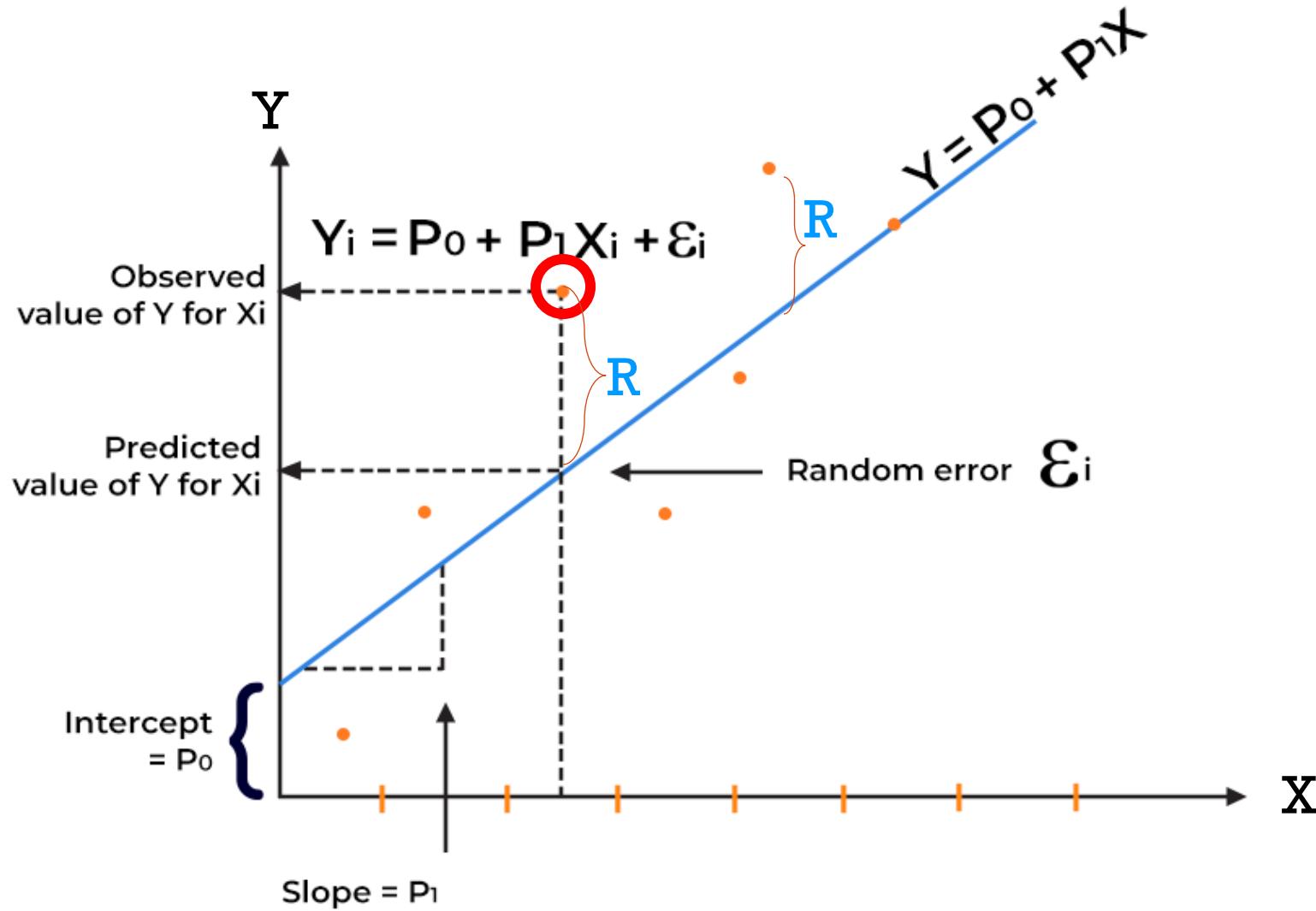
REGULARIZATION



- Addition of a penalty term to the objective function .
- shrinks the coefficient estimates toward zero and possibly forces some of them to be exactly zero.



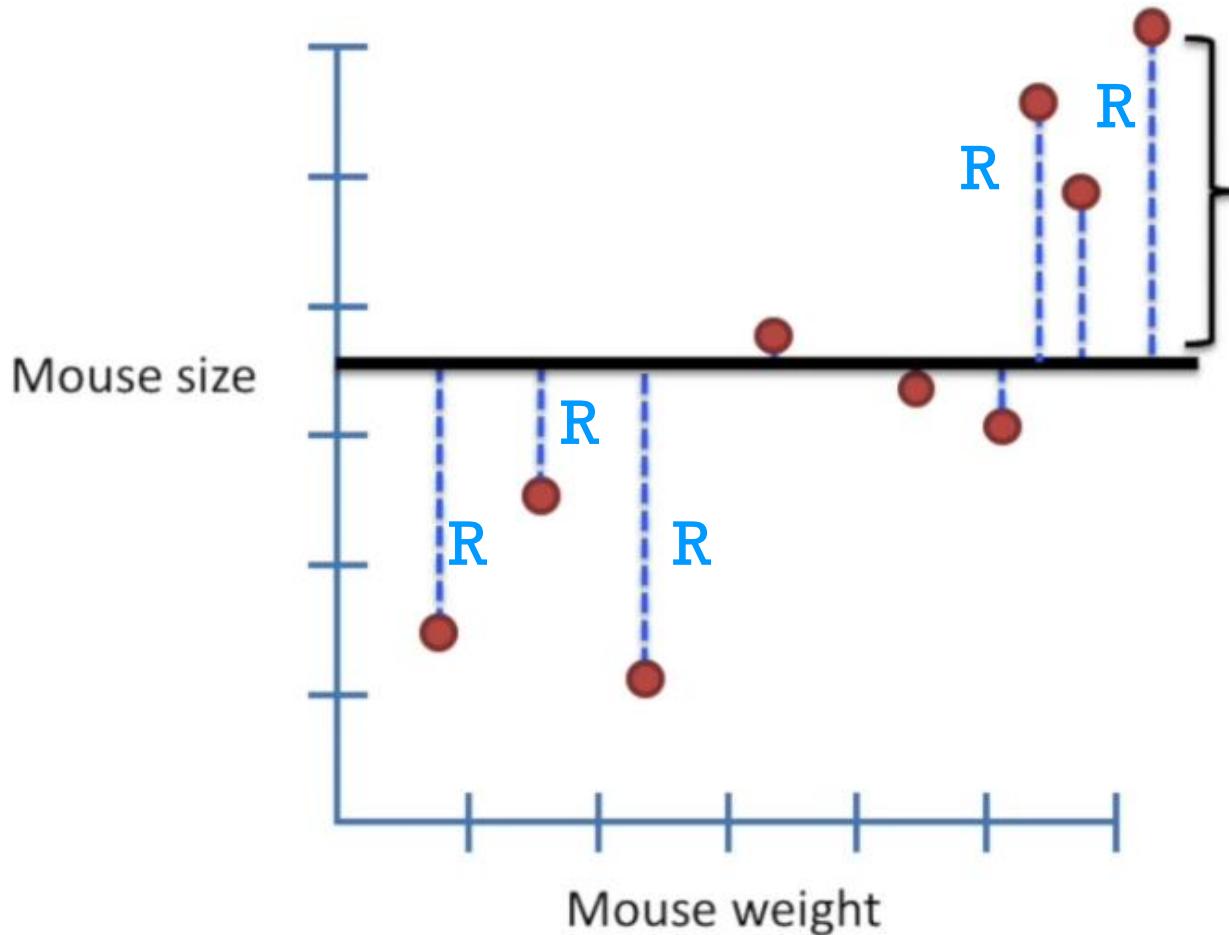
REGRESSION





L.B.B

Laboratory of Systems Biology
and Bioinformatics



$$Err = \sum_{k=1}^n R^2$$

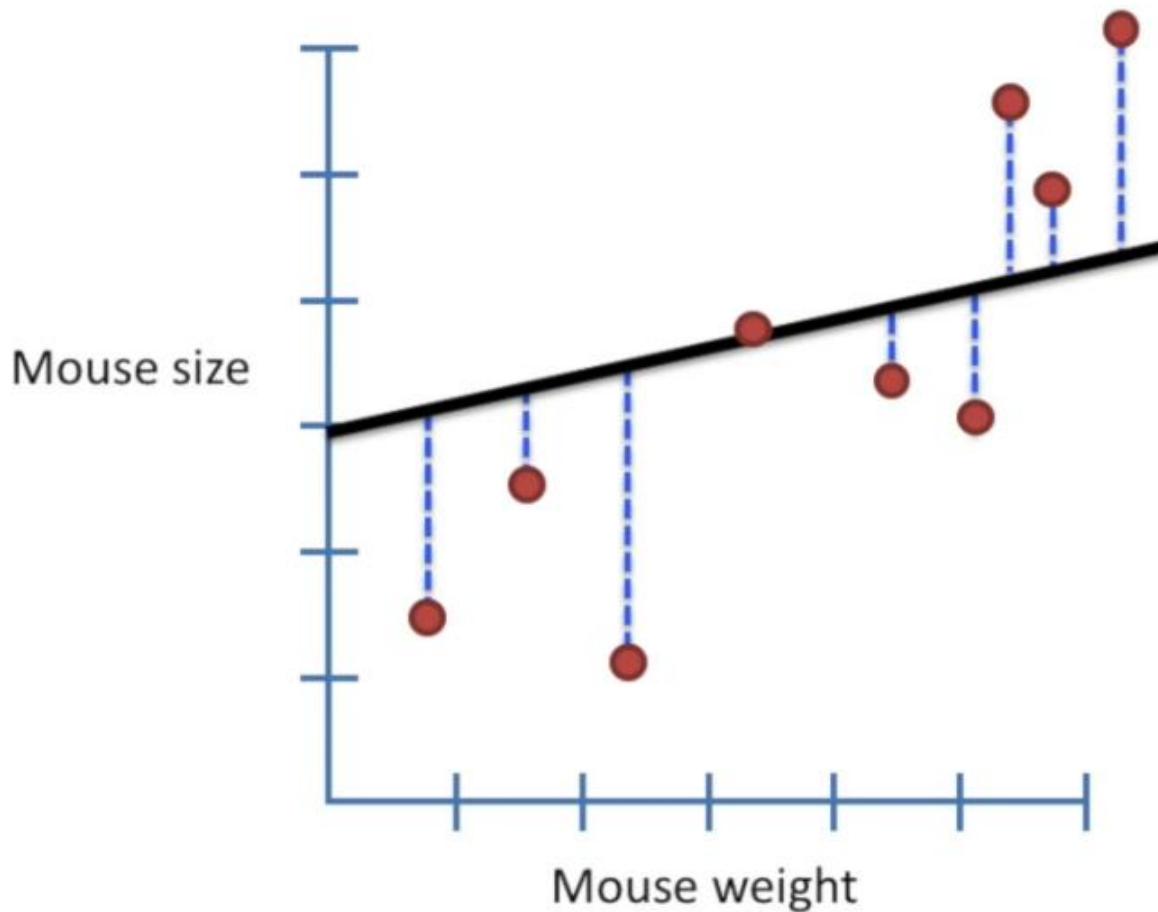
Second, measure the distance from the line to the data, square each distance, and then add them up.

Terminology alert!
The distance from a line to a data point is called a “**residual**”.



L.S.B.

Laboratory of Systems Biology
and Bioinformatics



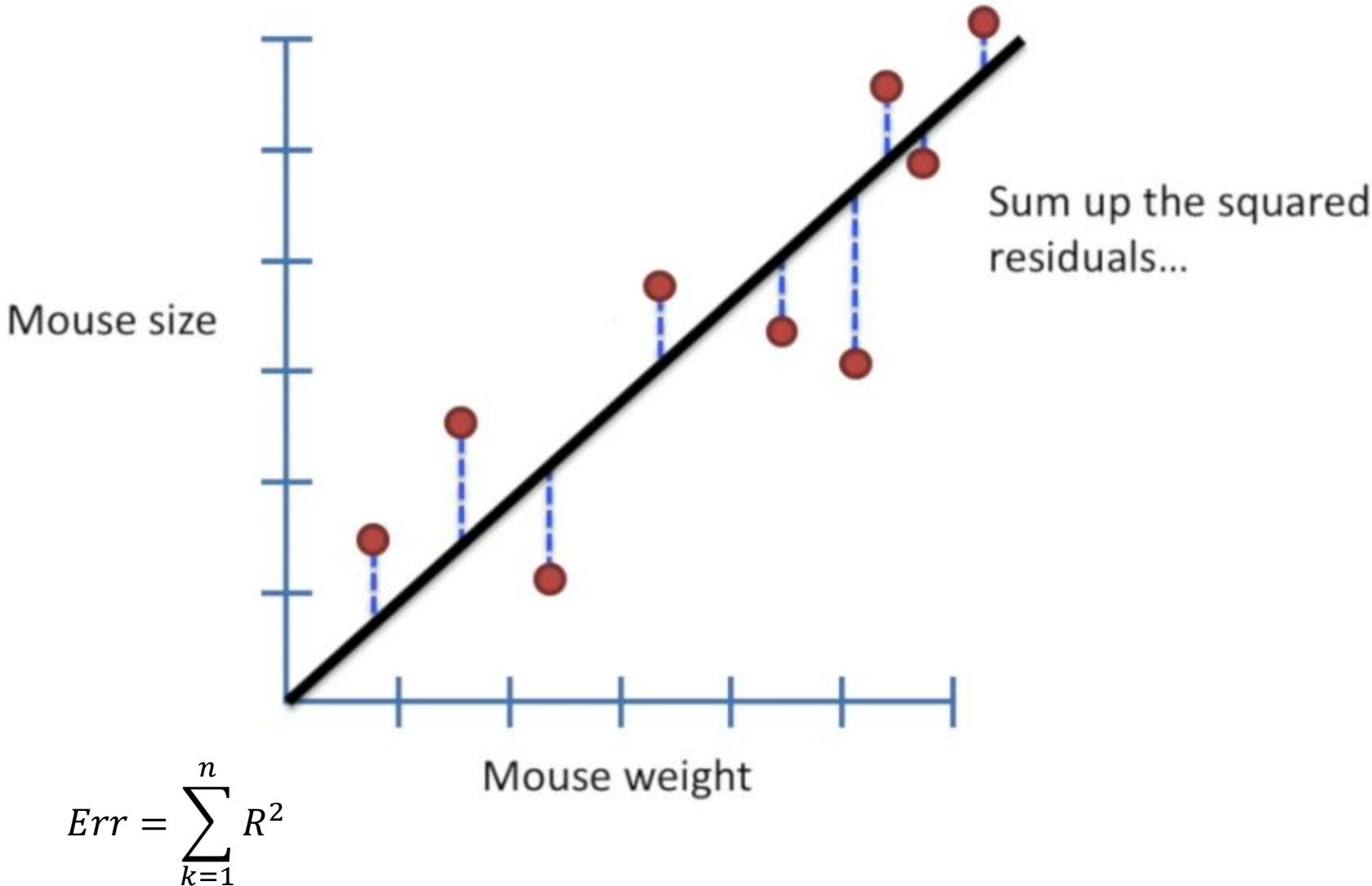
With the new line,
measure the
residuals, square
them, and then sum
up the squares.

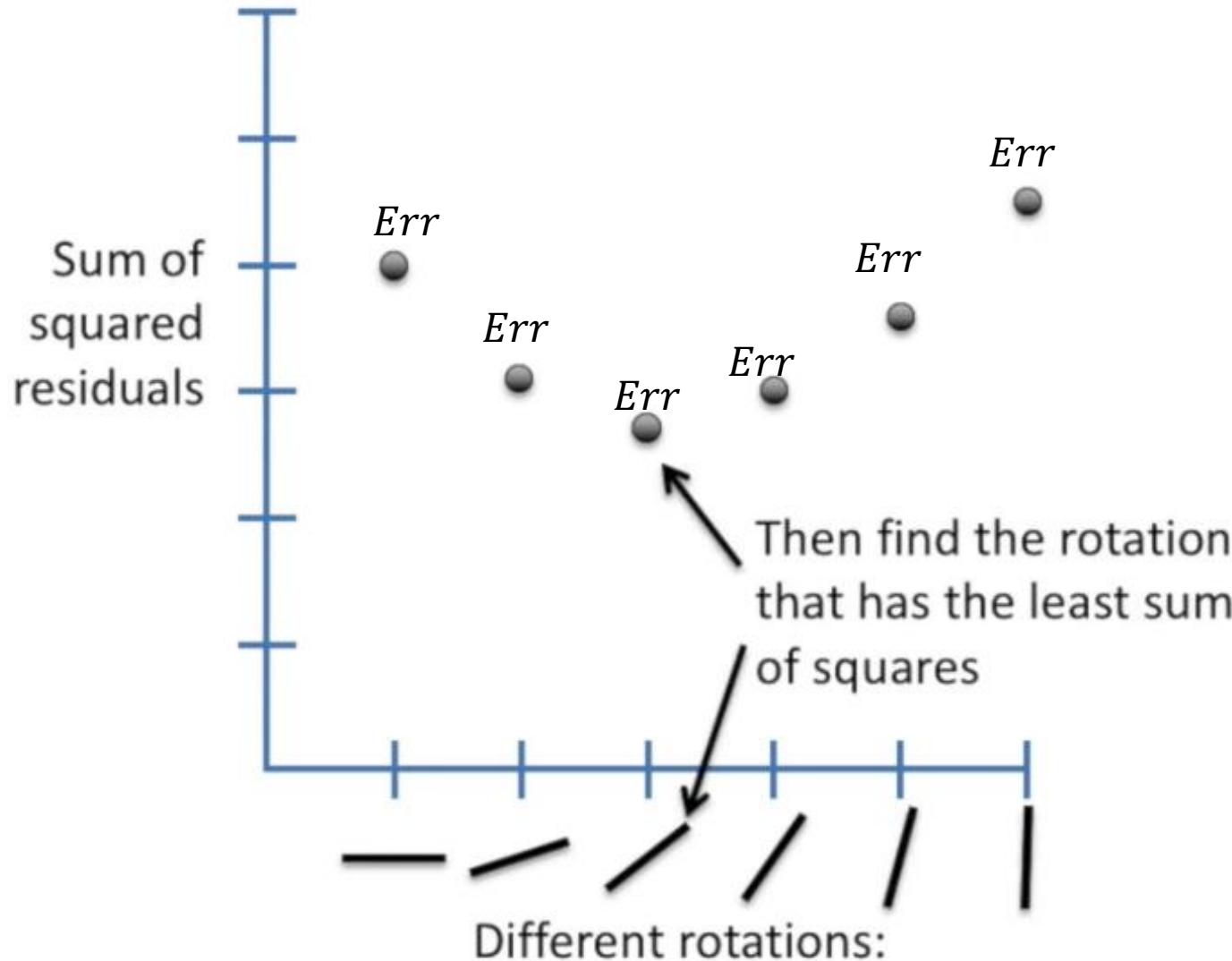
$$Err = \sum_{k=1}^n R^2$$



L.B.B

Laboratory of Systems Biology
and Bioinformatics

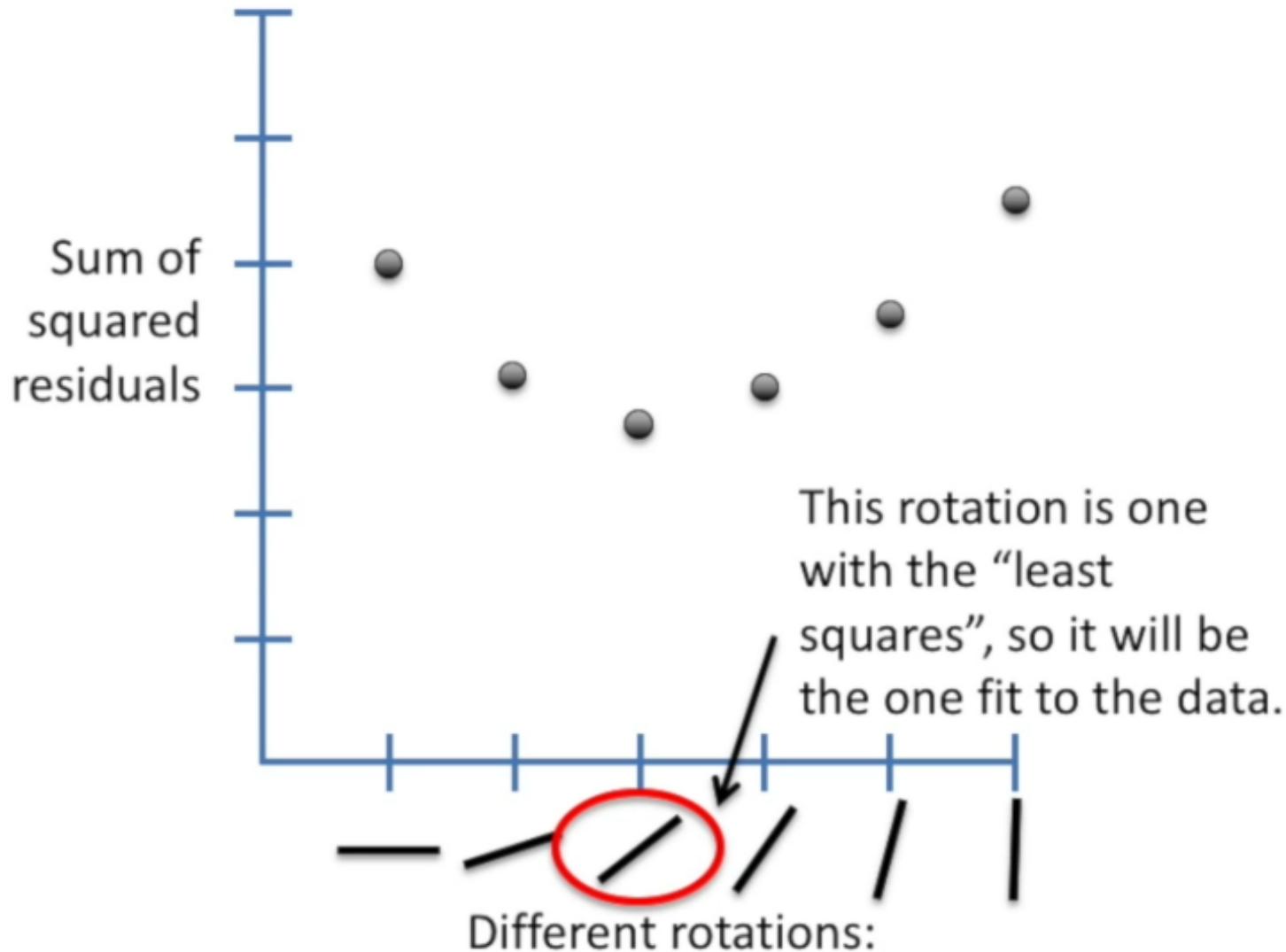


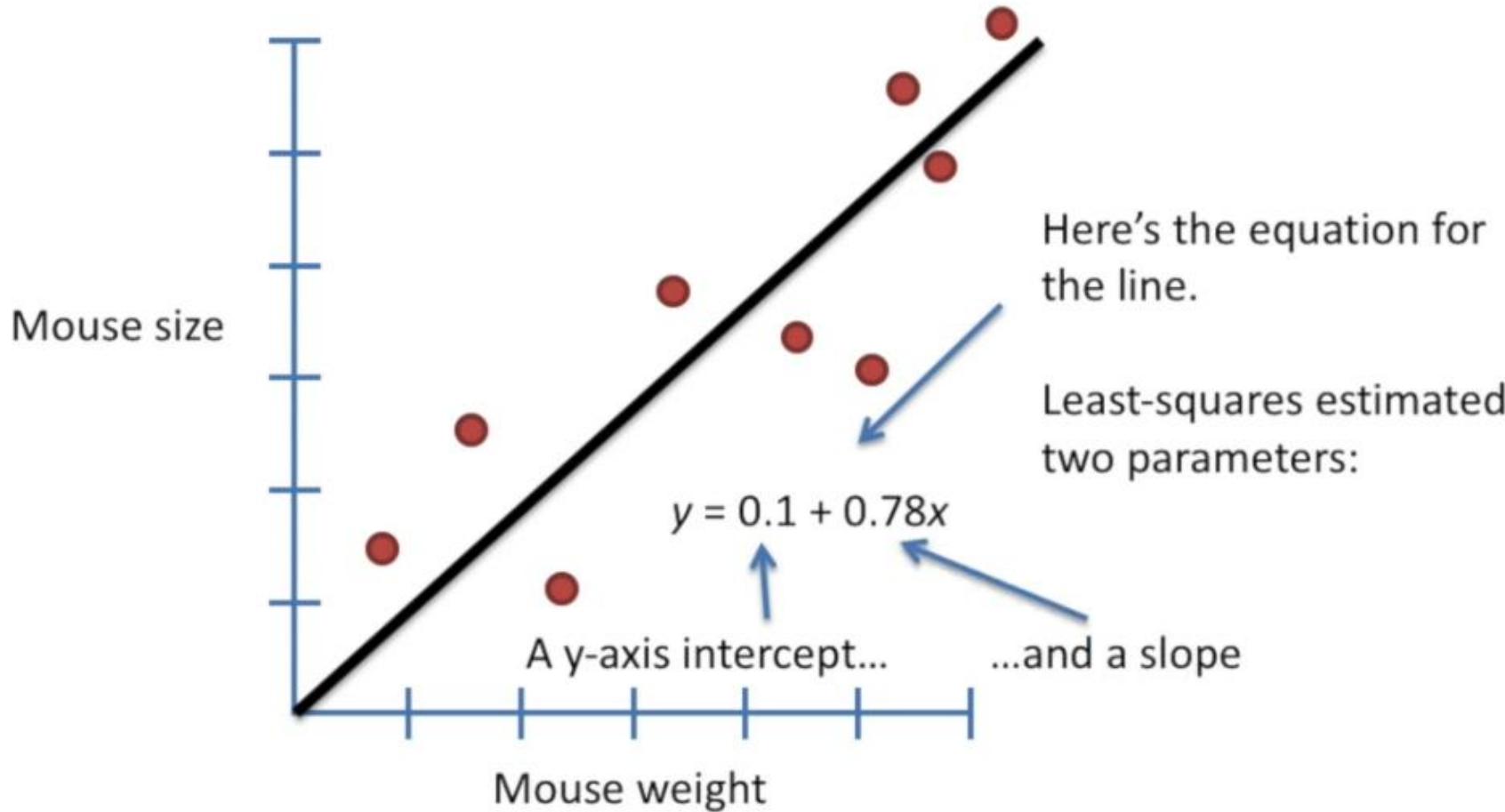




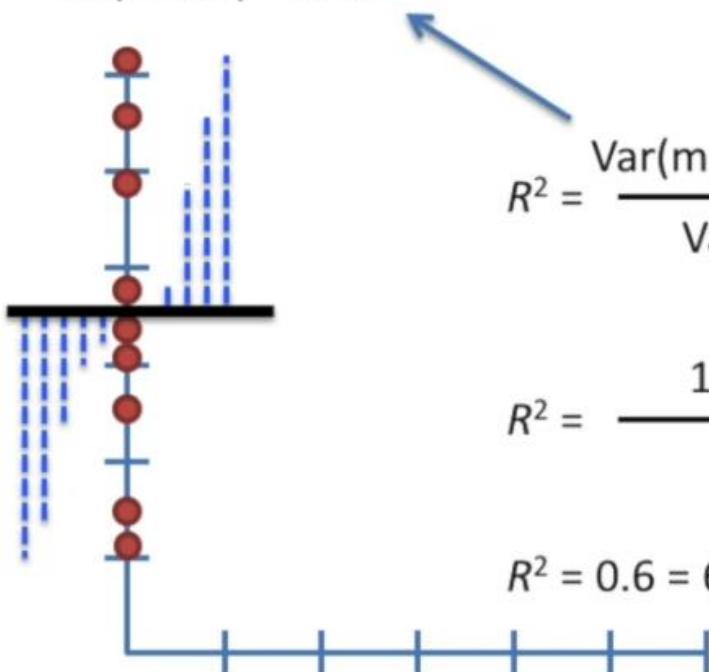
L.B.B

Laboratory of Systems Biology
and Bioinformatics





$\text{Var}(\text{mean}) = 11.1$



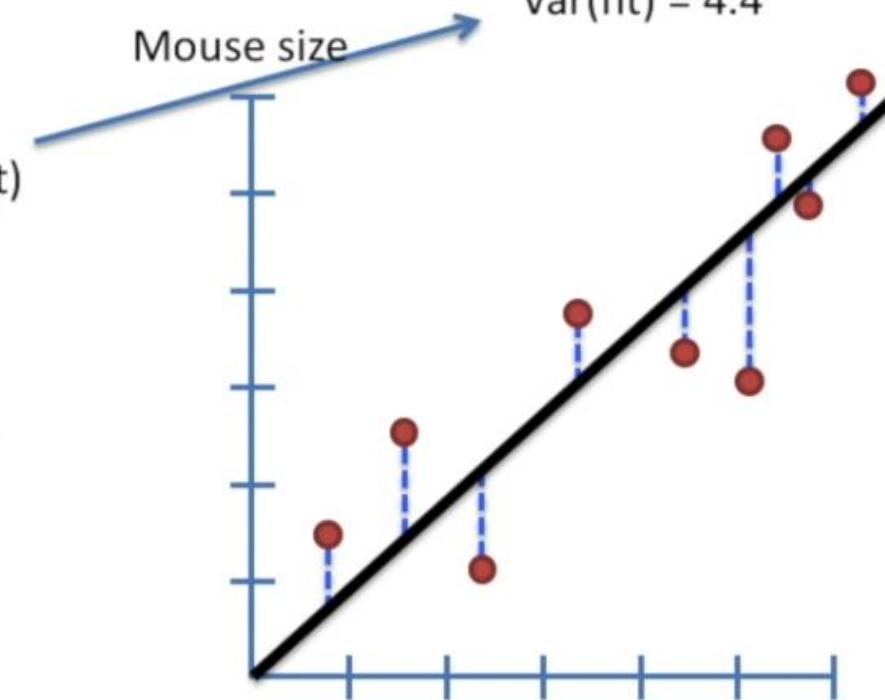
$$R^2 = \frac{\text{Var}(\text{mean}) - \text{Var}(\text{fit})}{\text{Var}(\text{mean})}$$

$$R^2 = \frac{11.1 - 4.4}{11.1}$$

$$R^2 = 0.6 = 60\%$$

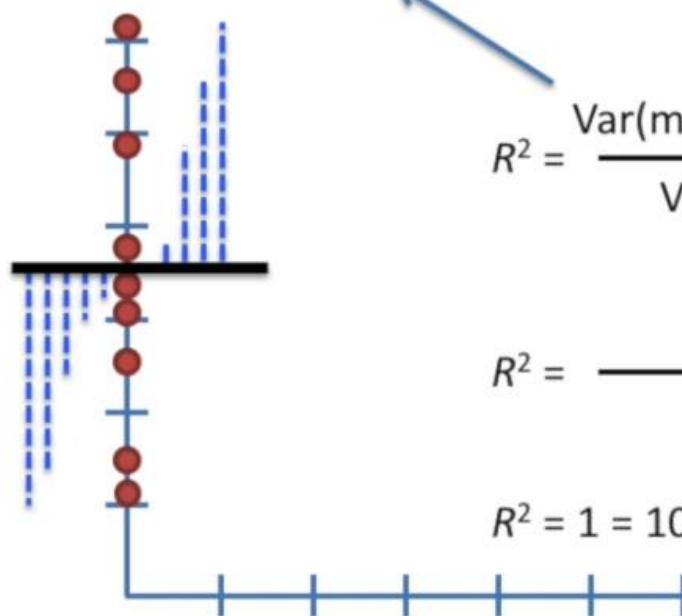
There is a 60% reduction in variance when we take the mouse weight into account.

$\text{Var}(\text{fit}) = 4.4$



Alternatively, we can say that mouse weight "explains" 60% of the variation in mouse size.

$\text{Var}(\text{mean}) = 11.1$

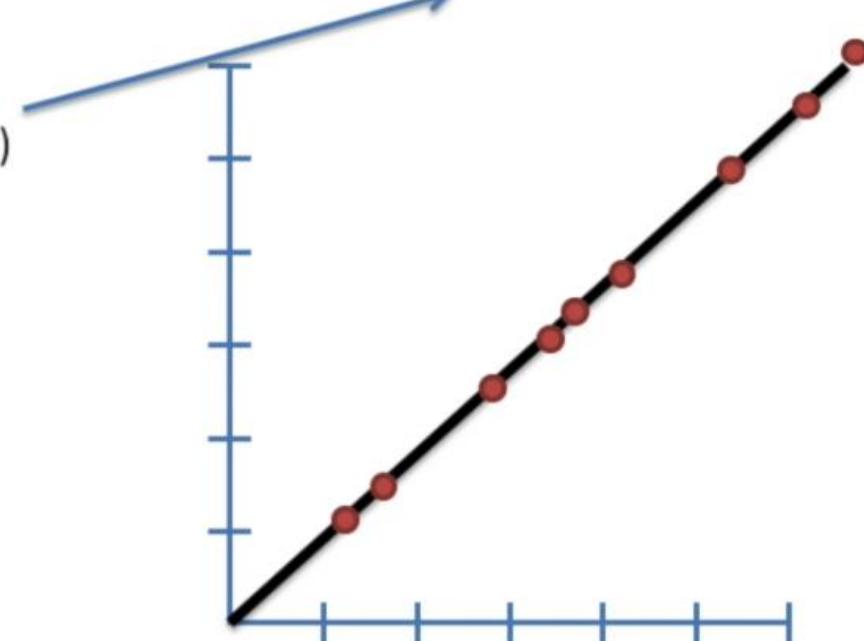


$$R^2 = \frac{\text{Var}(\text{mean}) - \text{Var}(\text{fit})}{\text{Var}(\text{mean})}$$

$$R^2 = \frac{11.1 - 0}{11.1}$$

$$R^2 = 1 = 100\%$$

$\text{Var}(\text{fit}) = 0$

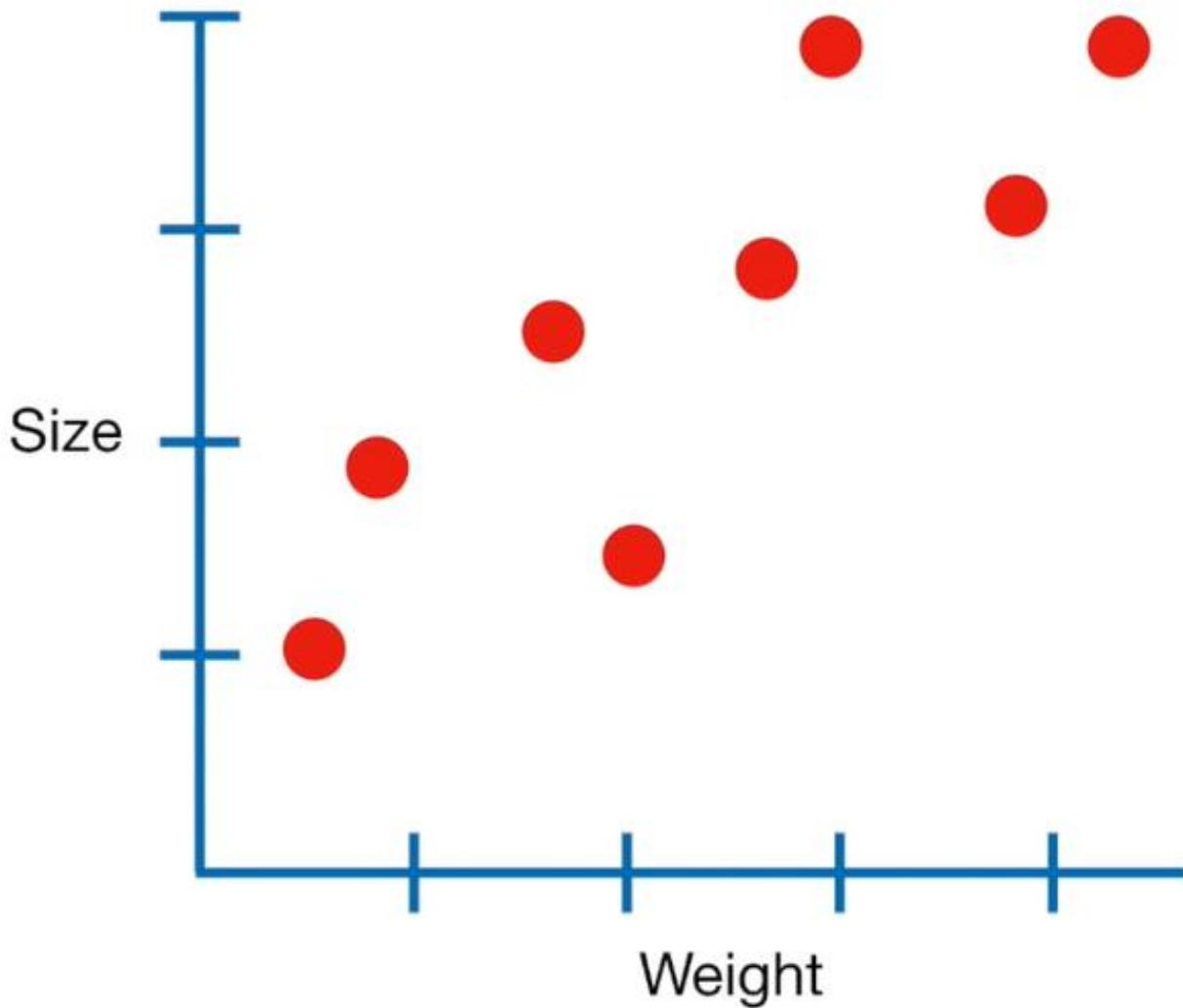


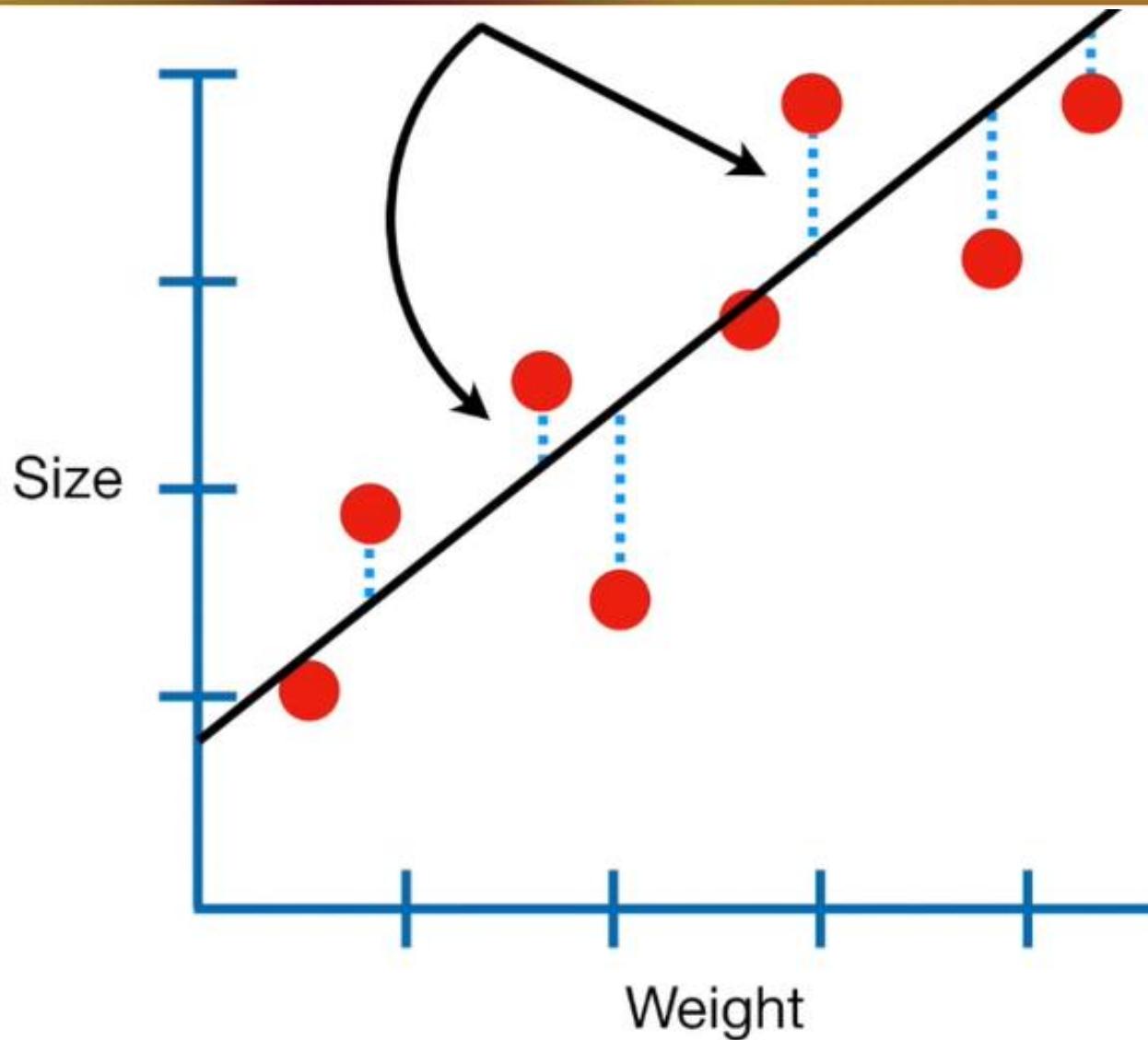
In this case, mouse weight “explains” 100% of the variation in mouse size.



L.B.B

Laboratory of Systems Biology
and Bioinformatics





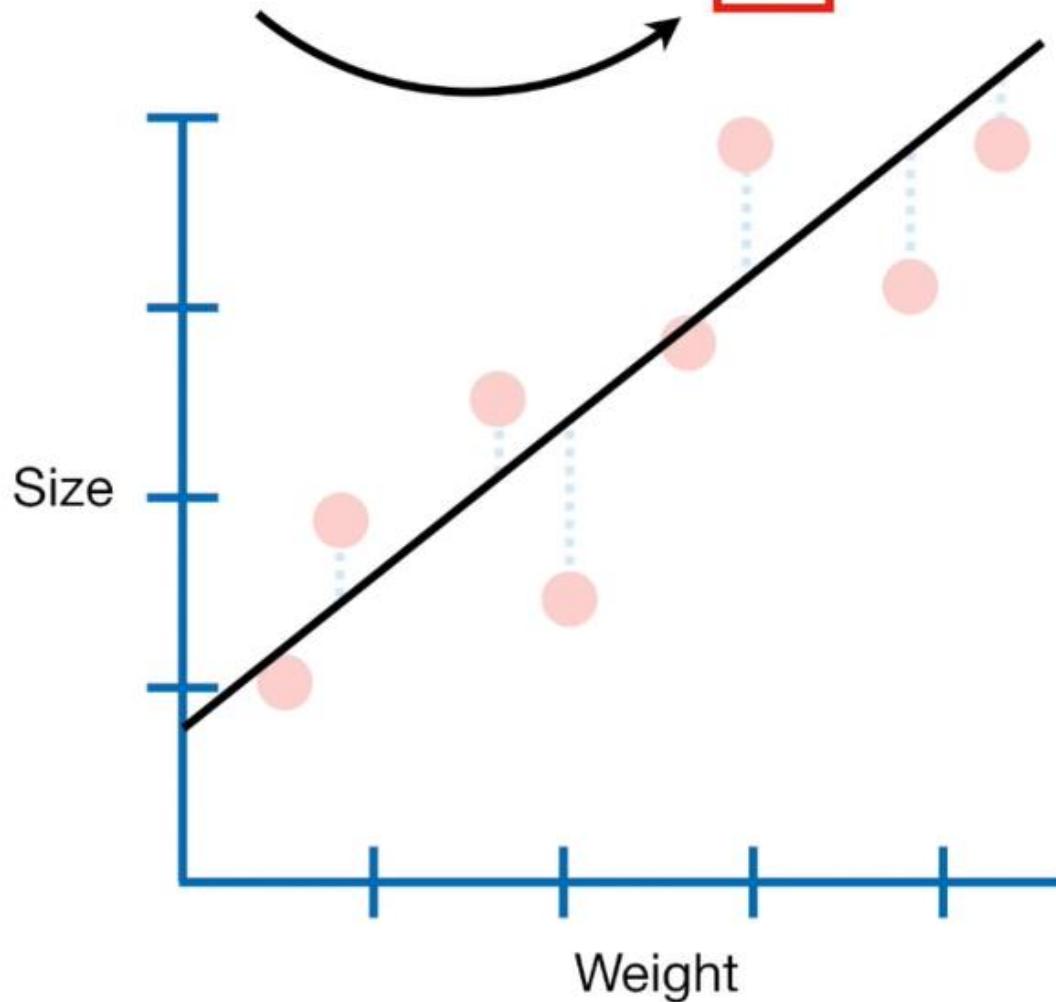


L.B.B

Laboratory of Systems Biology
and Bioinformatics

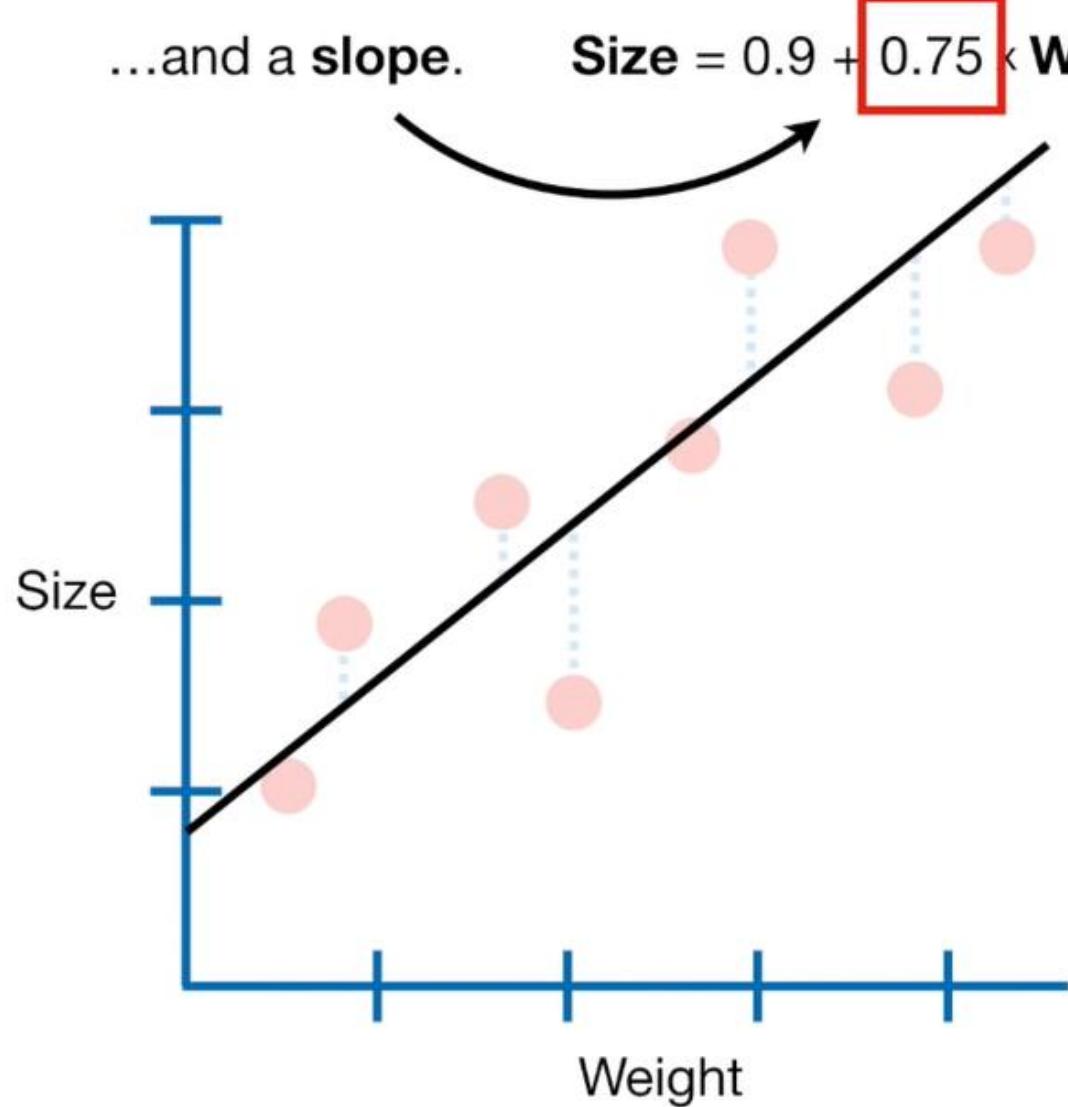
...a y-axis intercept...

$$\text{Size} = \boxed{0.9} + 0.75 \times \text{Weight}$$



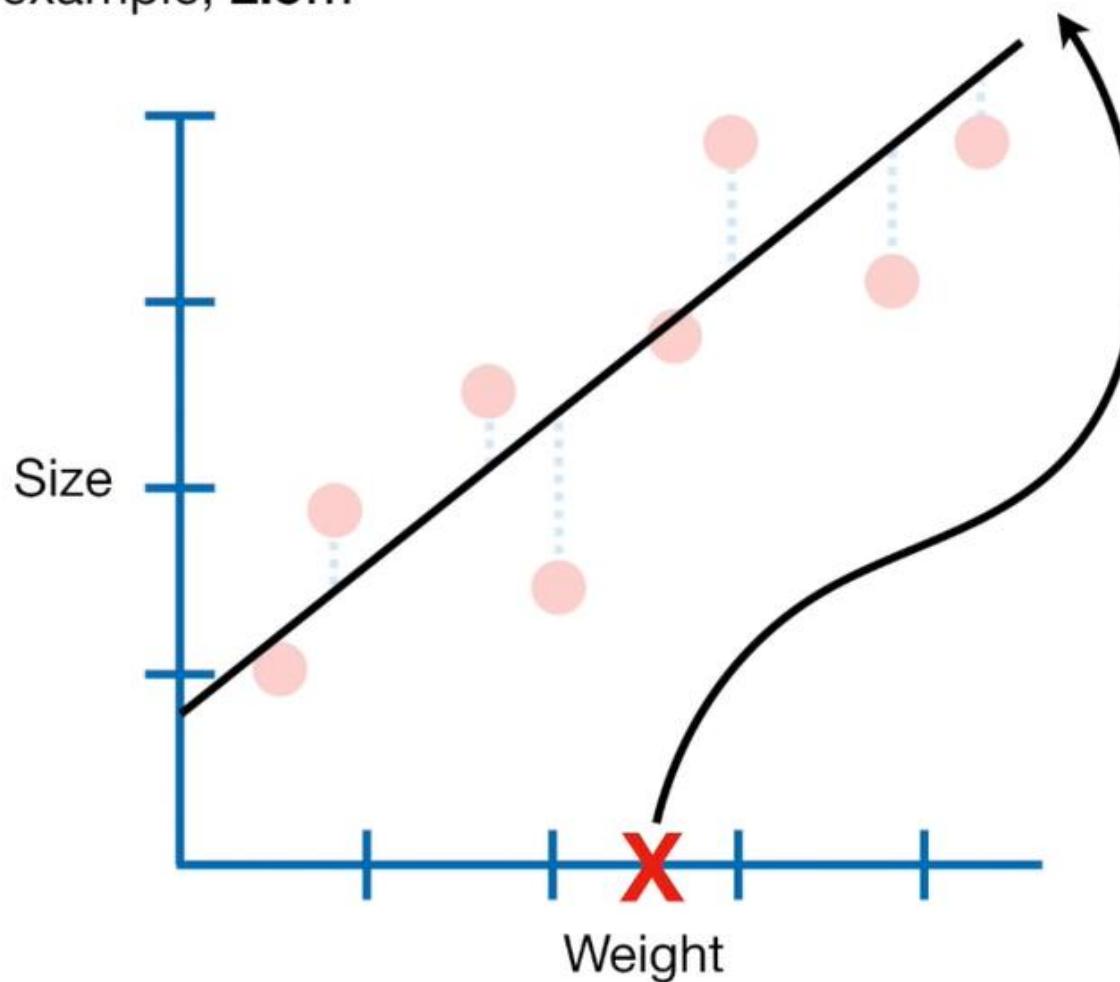
...and a **slope**.

$$\text{Size} = 0.9 + 0.75 \times \text{Weight}$$



We can plug in a value for **Weight**, for example, **2.5**...

$$\text{Size} = 0.9 + 0.75 \times \text{Weight}$$

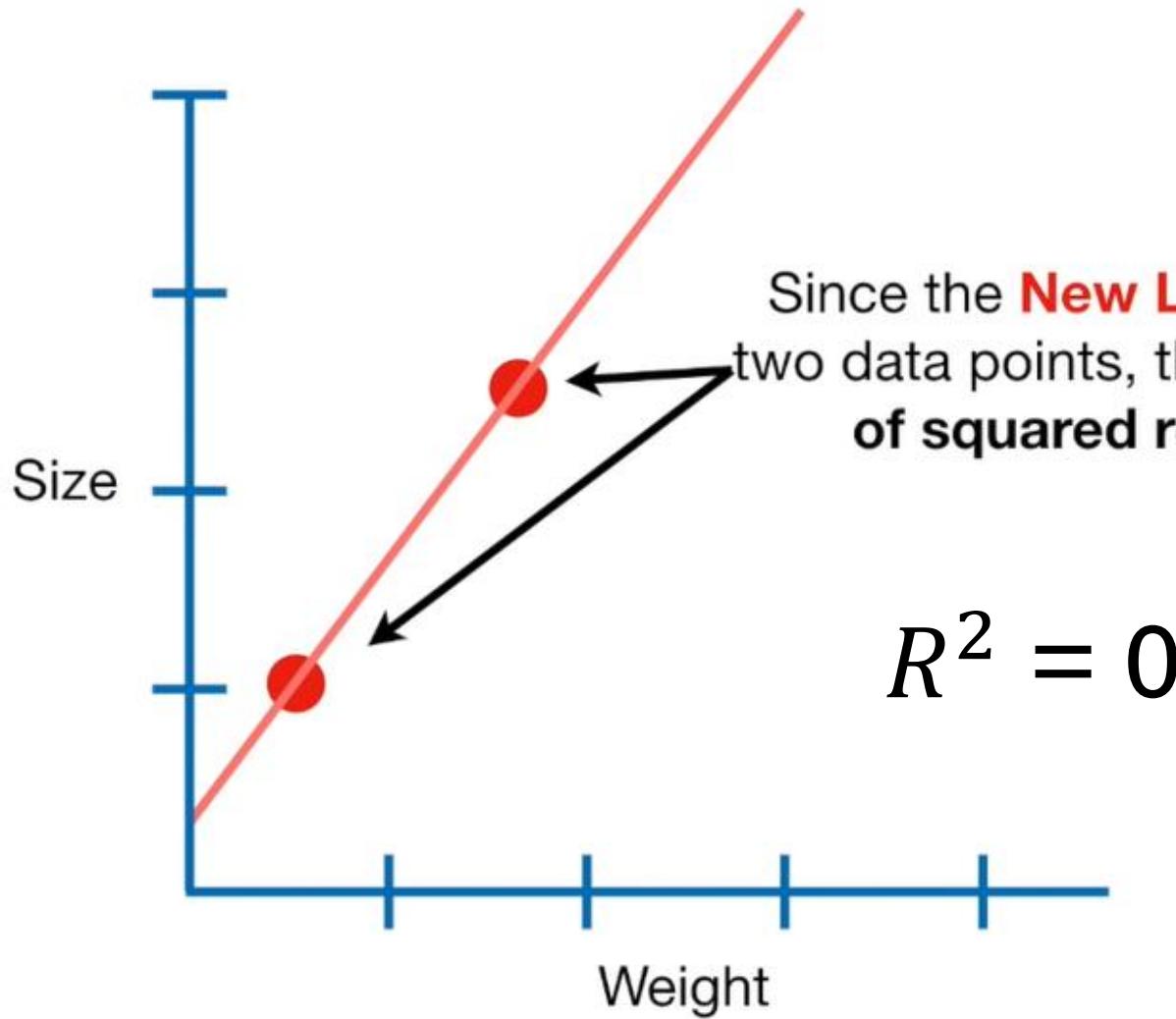




L.B.B

Laboratory of Systems Biology
and Bioinformatics

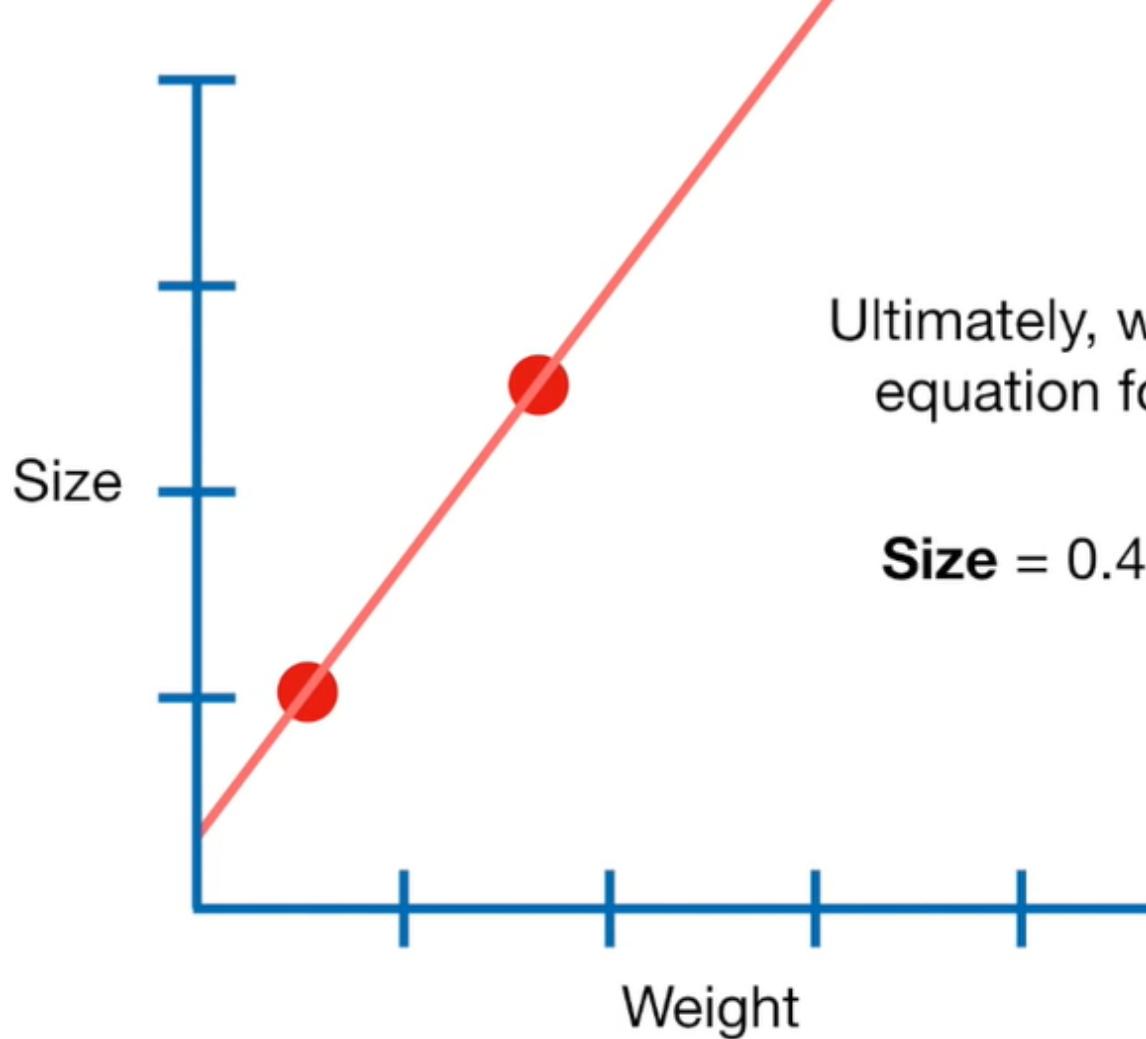
Just fit





L.B.B

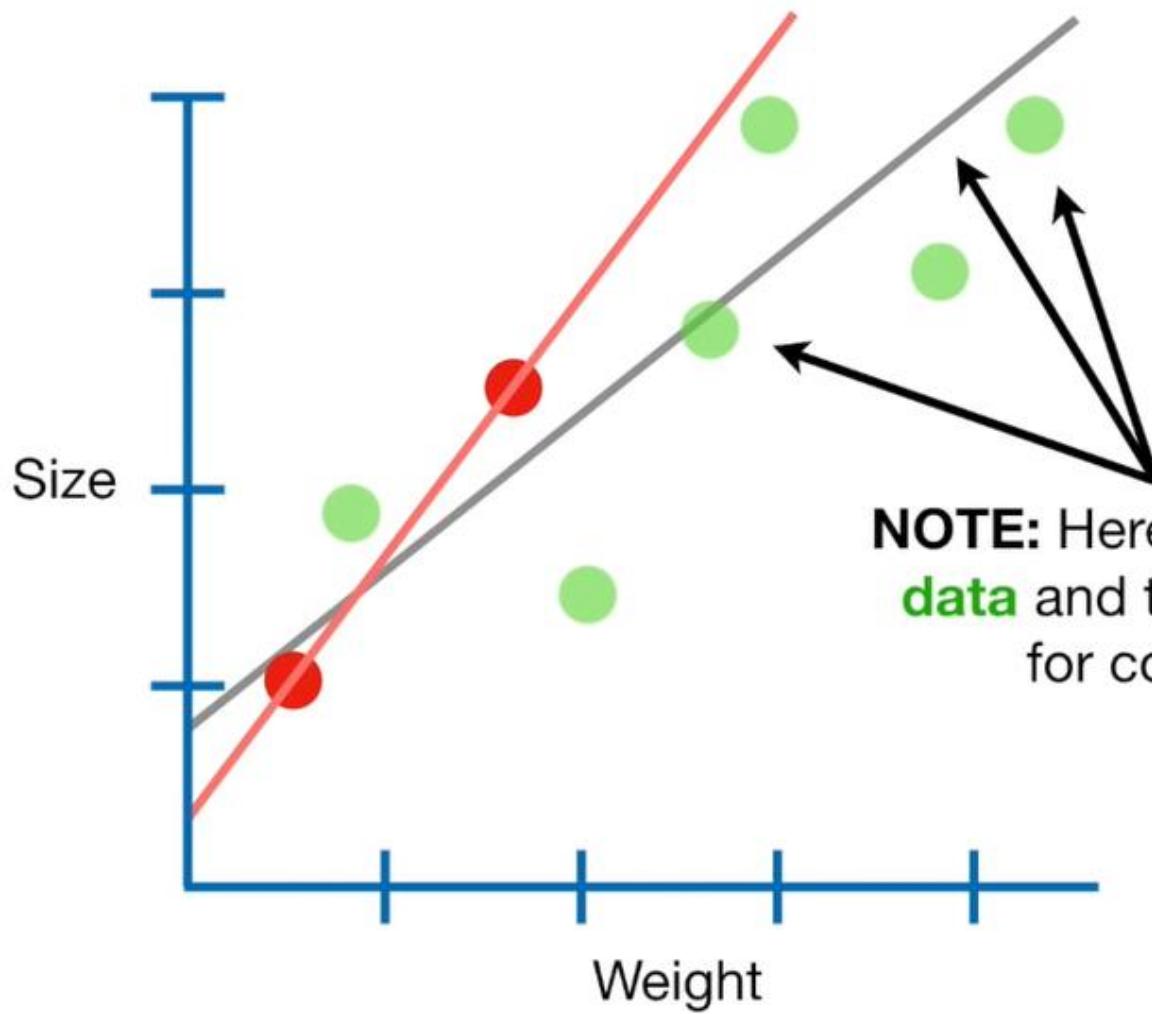
Laboratory of Systems Biology
and Bioinformatics



Ultimately, we end up with this equation for the **New Line**:

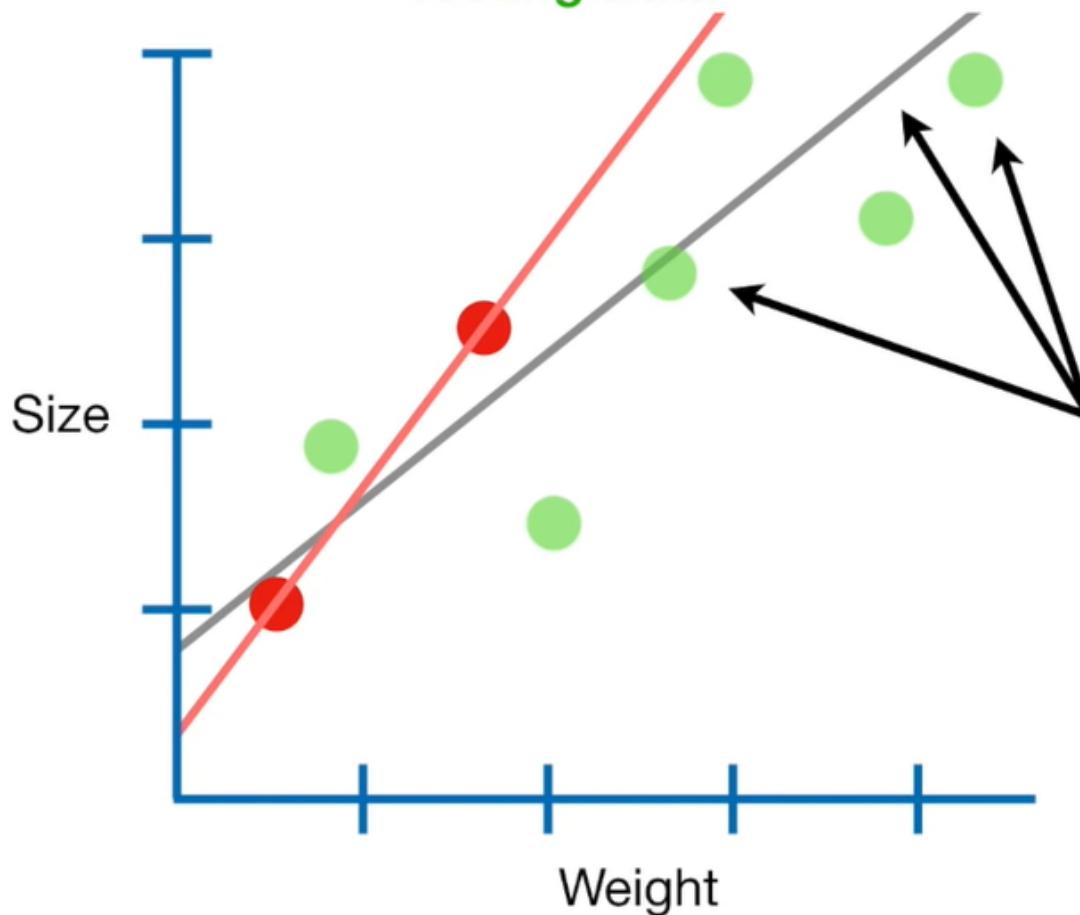
$$\text{Size} = 0.4 + 1.3 \times \text{Weight}$$

Model Fitting and Learning



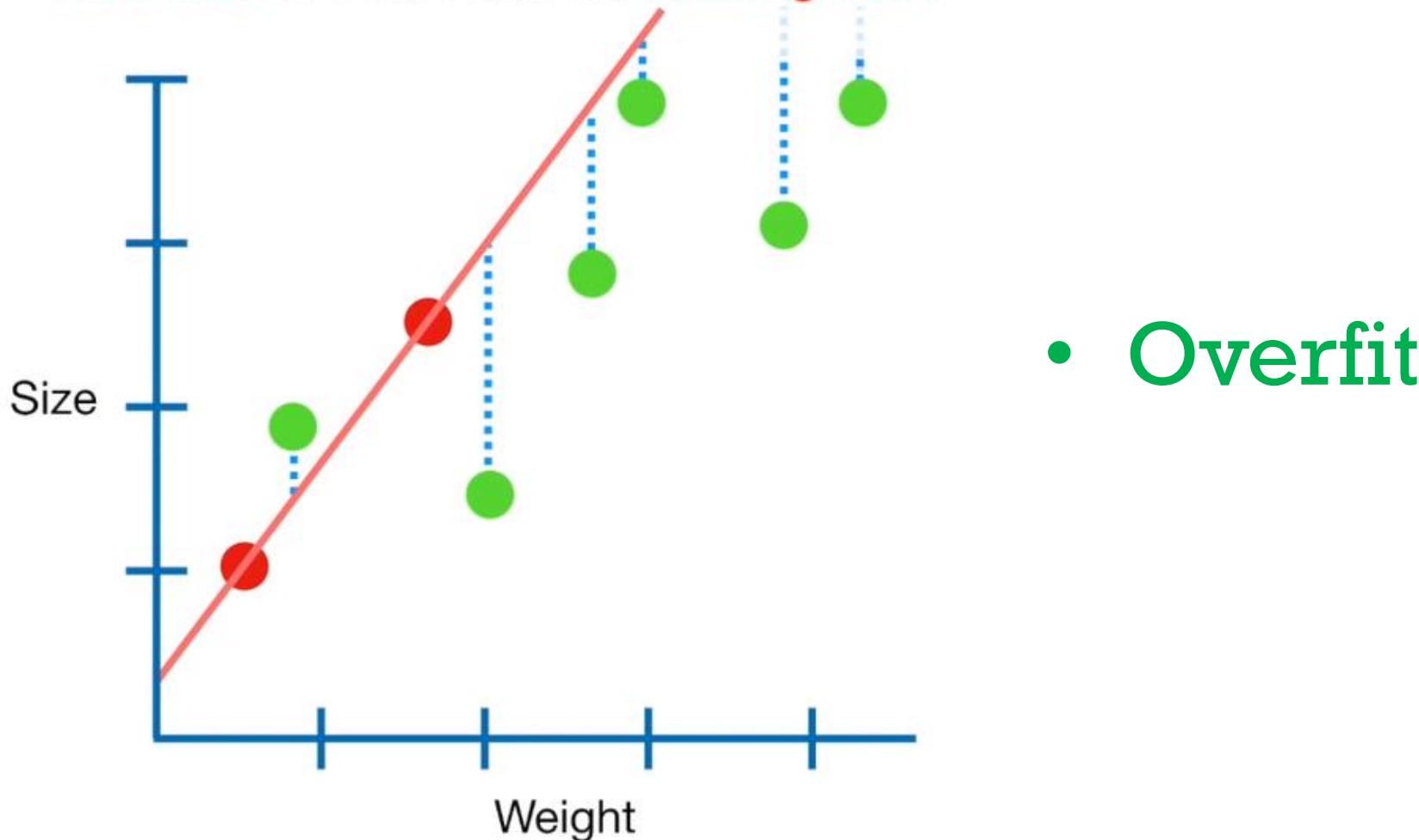
NOTE: Here are the **original data** and the **original line** for comparison.

Let's call the **Two Red Dots** the **Training Data**, and the remaining **Green Dots** the **Testing Data**.



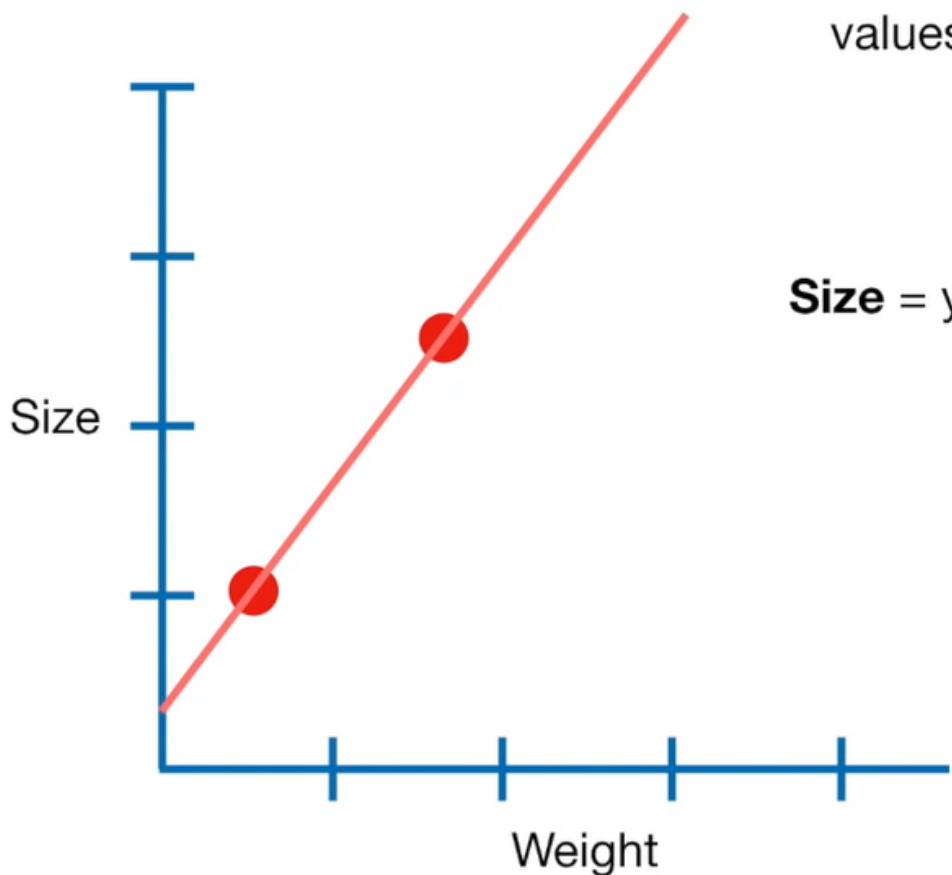
- **Training set**
- **Test Set**

In machine learning lingo, we'd say that the
New Line is **Over Fit** to the **Training Data**.



- Regularization
- Penalty
- ..

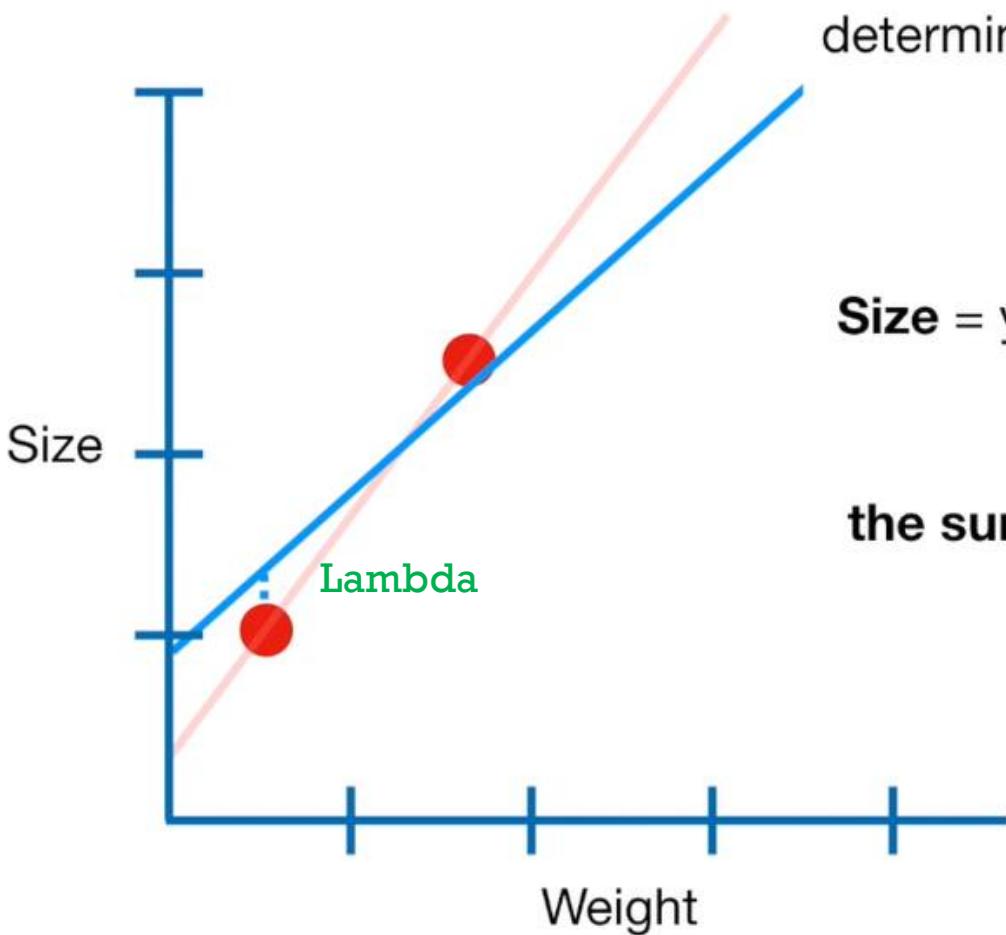
RIDGE (L2) REGRESSION



When **Least Squares** determines values for the parameters in this equation...



Size = y-axis intercept + slope \times **Weight**



In contrast, when **Ridge Regression** determines values for the parameters in this equation...



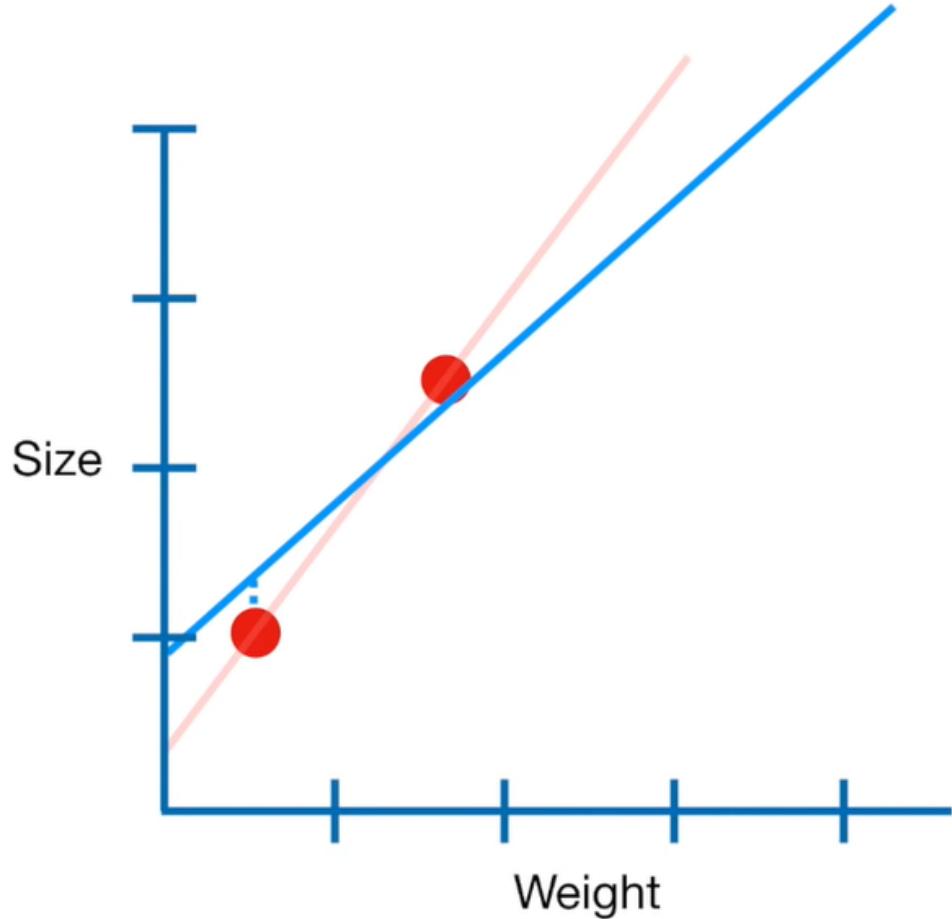
Size = y-axis intercept + slope \times **Weight**

...it minimizes...

the sum of the squared residuals

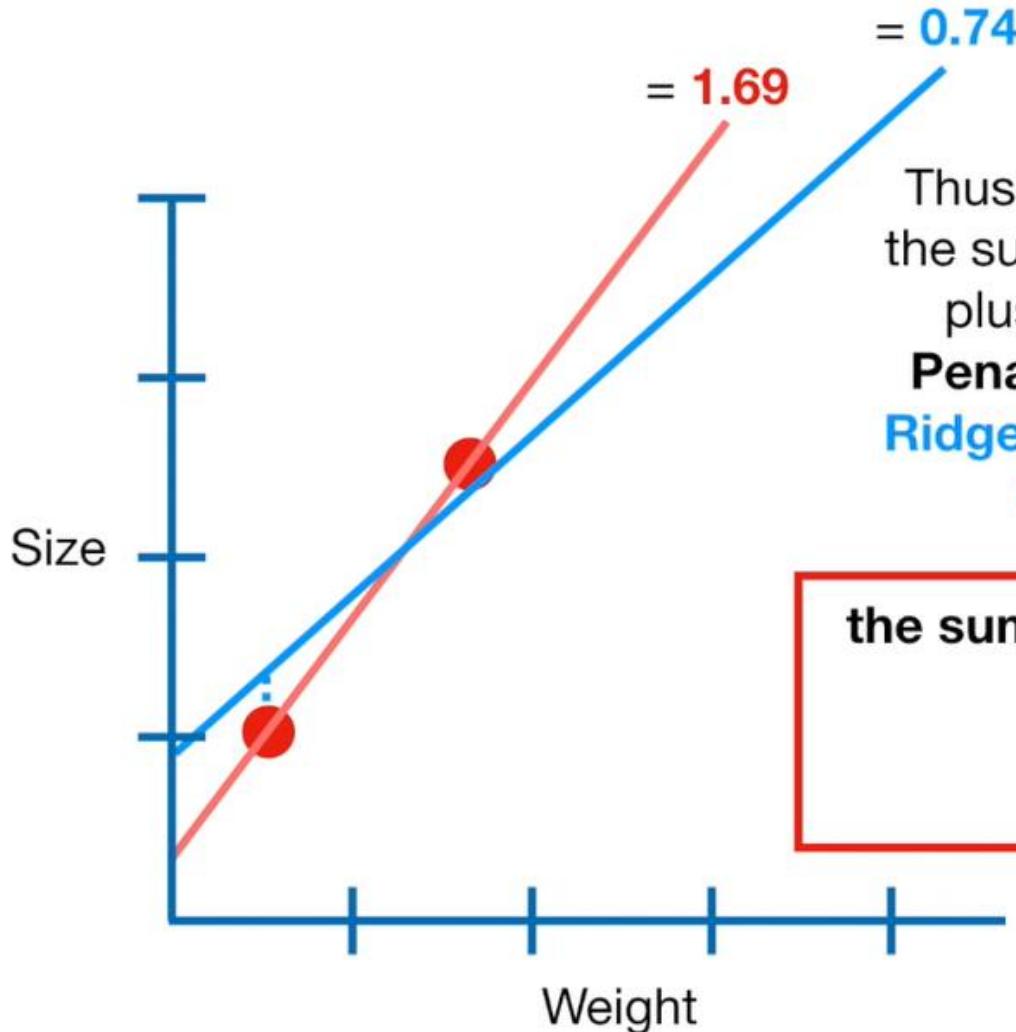
+

$\lambda \times \text{the slope}^2$



...and when we do
the math we get...

$$\begin{aligned} & 0.3^2 + 0.1^2 \\ & + \quad = 0.09 + 0.01 + 0.64 \\ & 1 \times 0.8^2 \end{aligned}$$



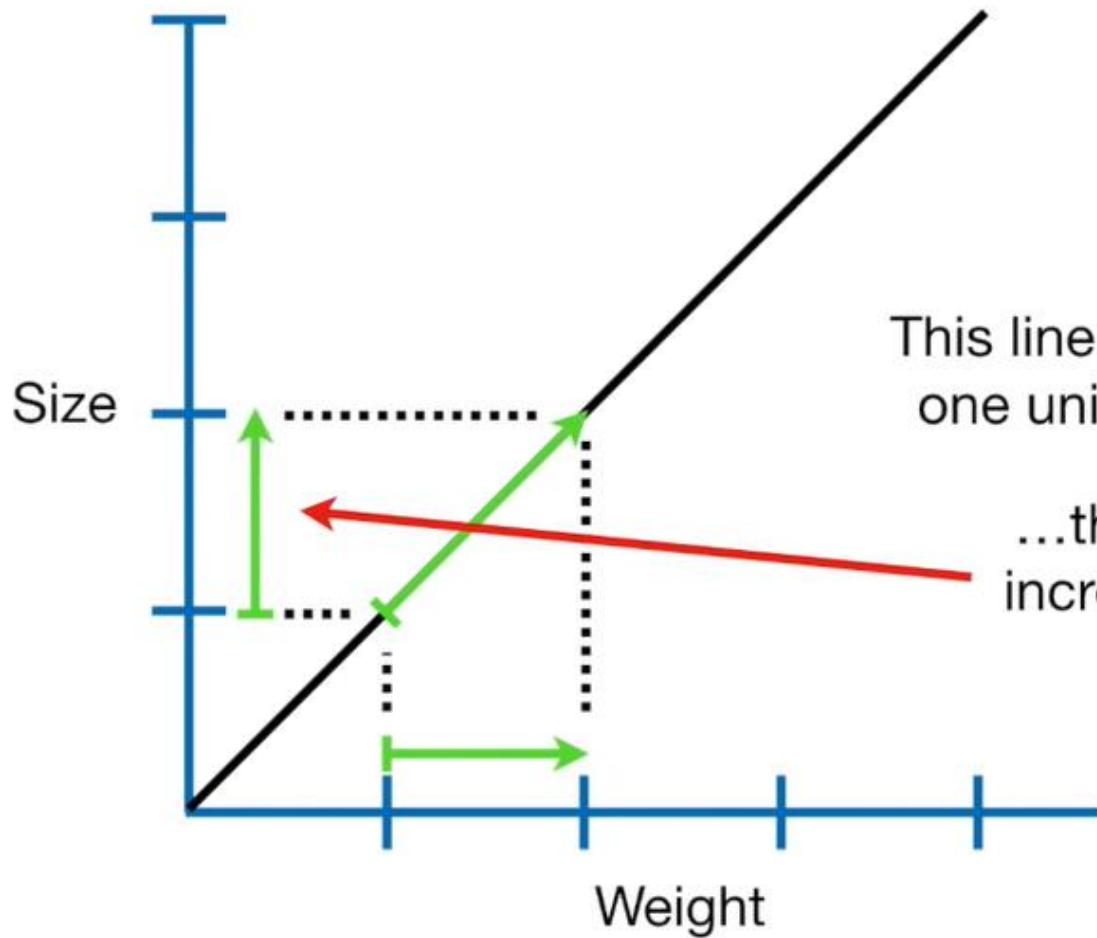
Thus, if we wanted to minimize the sum of the squared residuals plus the **Ridge Regression Penalty**, we would choose the **Ridge Regression Line** over the **Least Squares Line**.

the sum of the squared residuals
+
 $\lambda \times \text{the slope}^2$



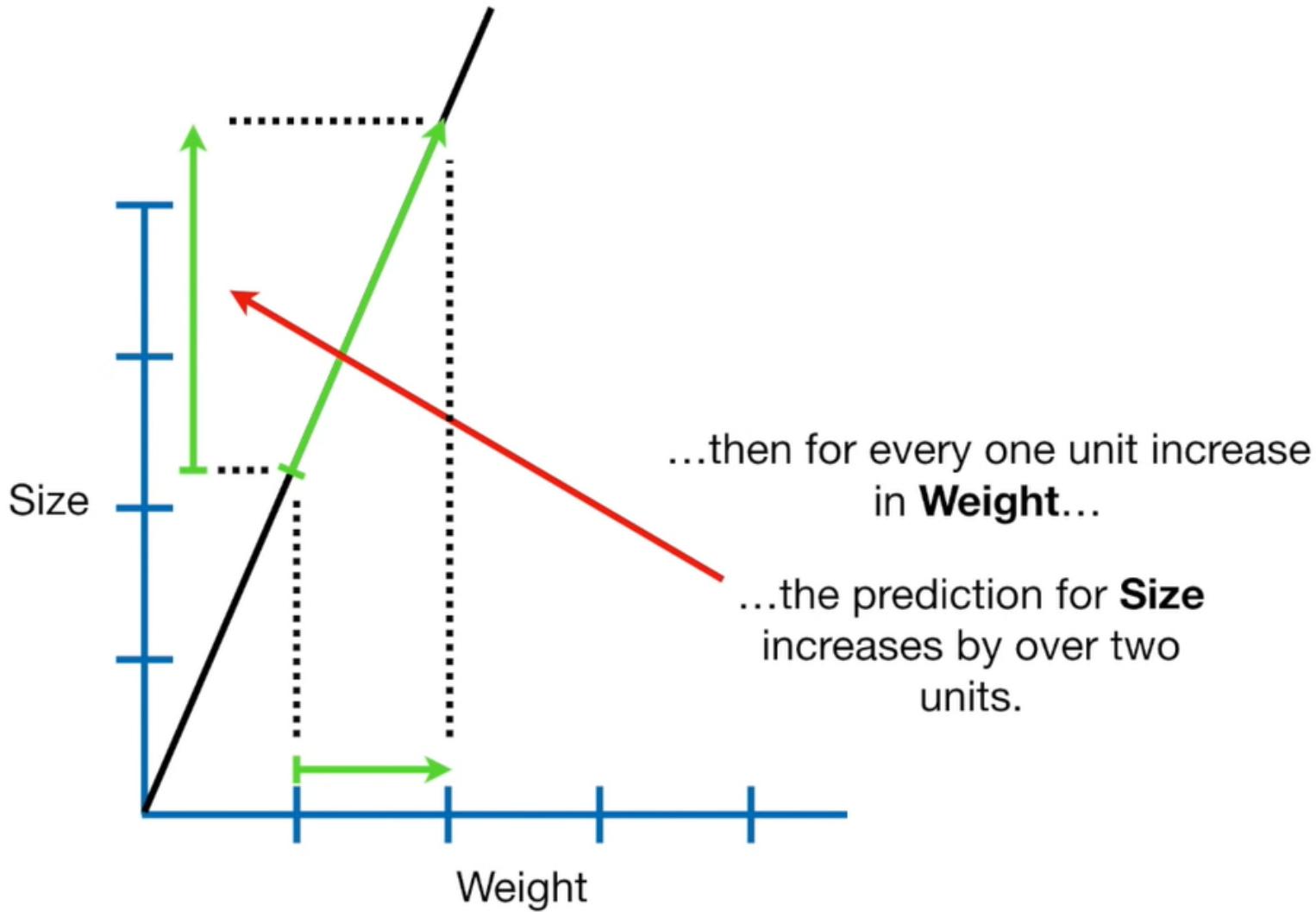
L.B.B

Laboratory of Systems Biology
and Bioinformatics

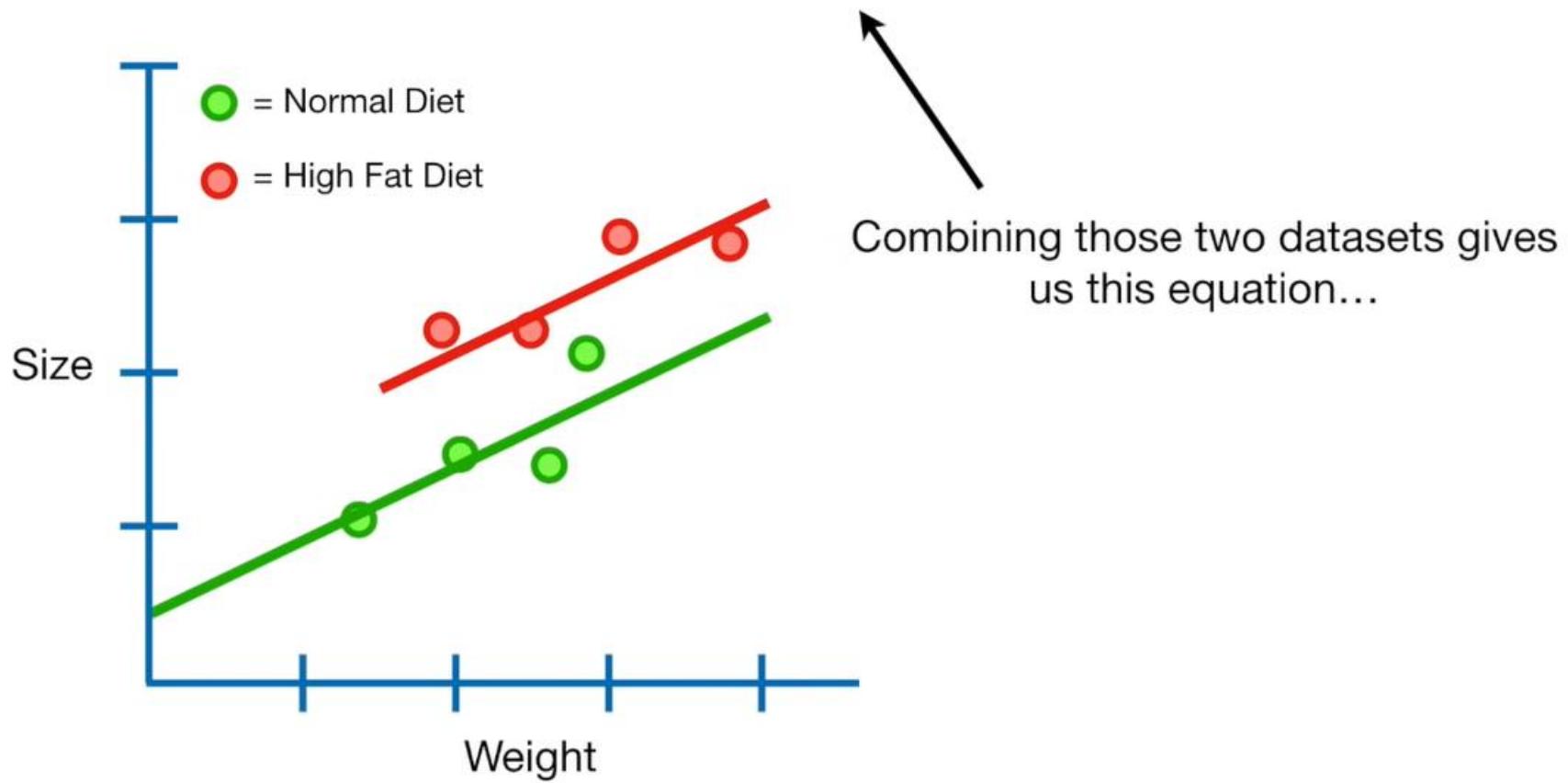


This line suggests that for every one unit increase in **Weight**...

...there is a one unit increase in predicted **Size**.



Size = y-intercept + slope \times **Weight** + diet difference \times **High Fat Diet**



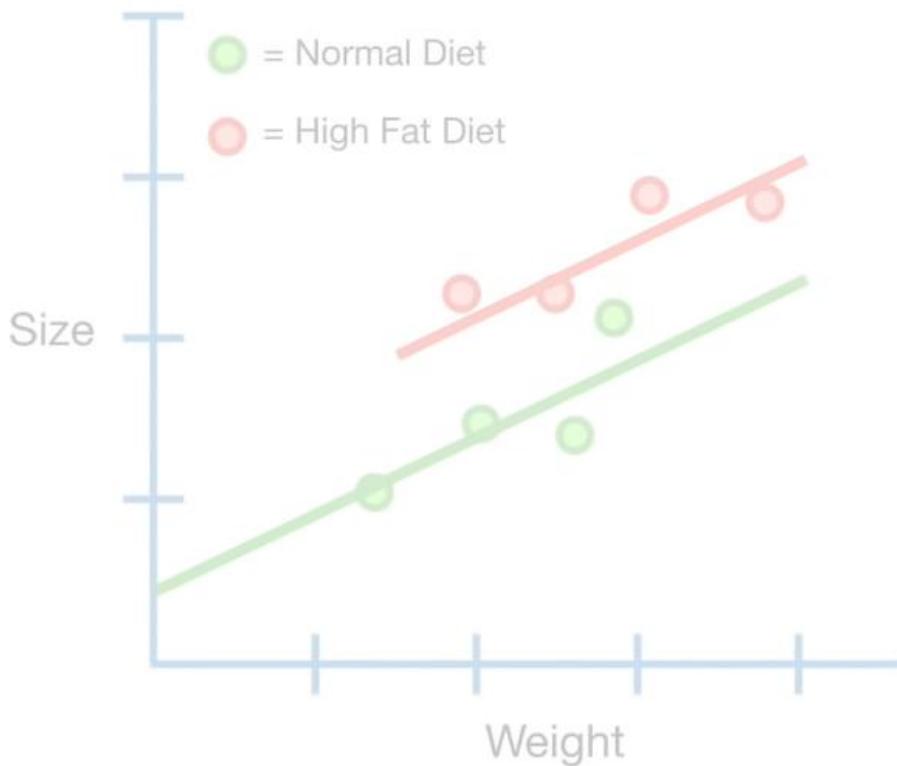
Combining those two datasets gives us this equation...



L.B.B

Laboratory of Systems Biology
and Bioinformatics

$$\text{Size} = \text{y-intercept} + \text{slope} \times \text{Weight} + \text{diet difference} \times \text{High Fat Diet}$$



Now the **Ridge Regression Penalty** contains the parameters for the **slope** and the **difference between diets**.

the sum of the squared residuals

$$\lambda \times (\text{slope}^2 + \text{diet difference}^2)$$

Size = y-intercept + slope × **Weight** + diet difference × **High Fat Diet**
+ astrological offset × **Sign** + airspeed scalar × **Airspeed of Swallow...**

Then the **Ridge Regression Penalty** would have all those parameters squared, except for the **y-intercept**.

$\lambda (\text{slope}^2 + \text{diet difference}^2 + \text{astrological offset}^2 + \text{airspeed scalar}^2 + \dots)$

Size = y-intercept + slope1 × **Gene1** + slope2 × **Gene2** + ... + slope10000 × **Gene10000**

For example, we might use gene expression measurements from 10,000 genes to predict **Size**...

...and that would mean we would need gene expression measurements from 10,001 mice.

Size = $y\text{-intercept} + \text{slope1} \times \mathbf{Gene1} + \text{slope2} \times \mathbf{Gene2} + \dots + \text{slope10000} \times \mathbf{Gene10000}$

It turns out that by adding the **Ridge Regression Penalty**, we can solve for all 10,001 parameters with only 500 (or even fewer) samples.

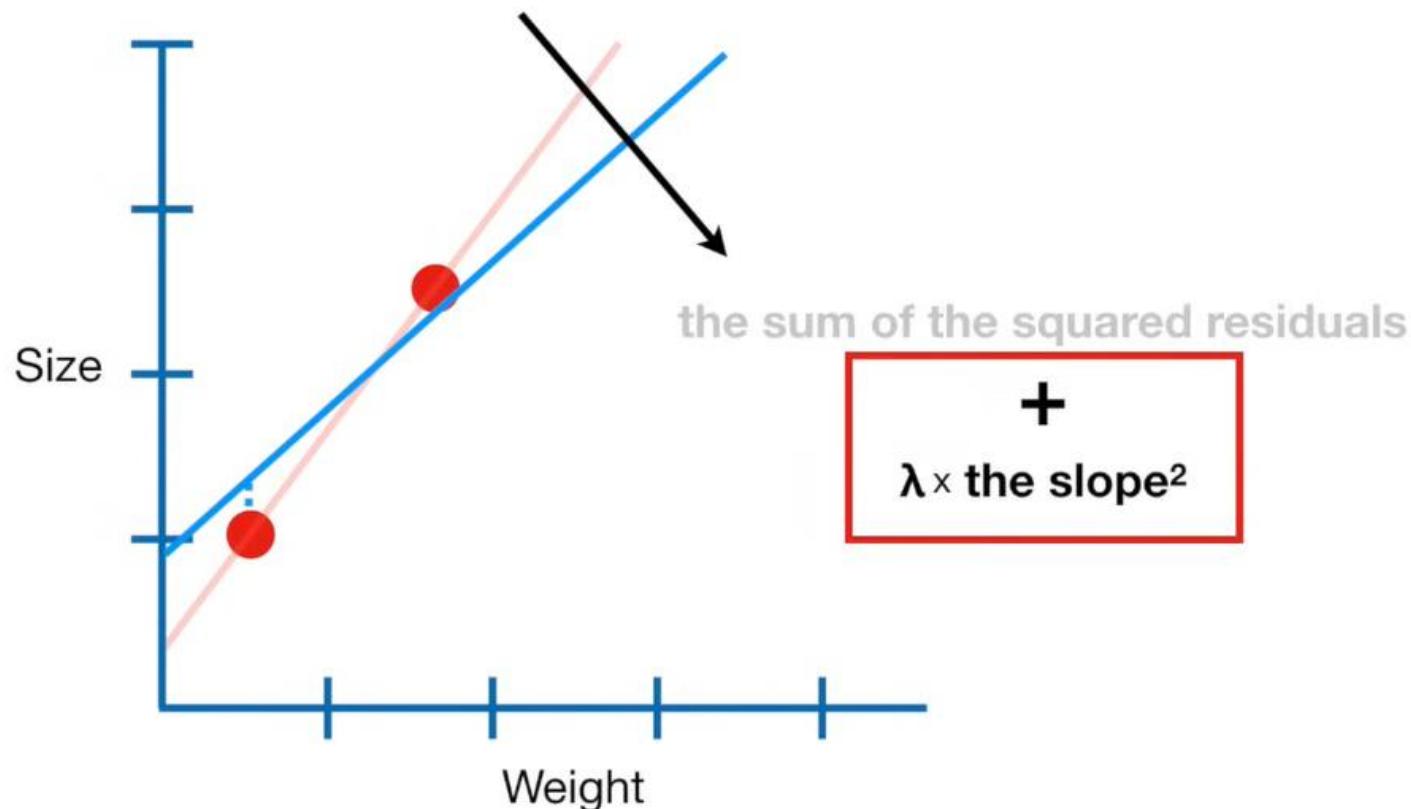
the sum of the squared residuals

+

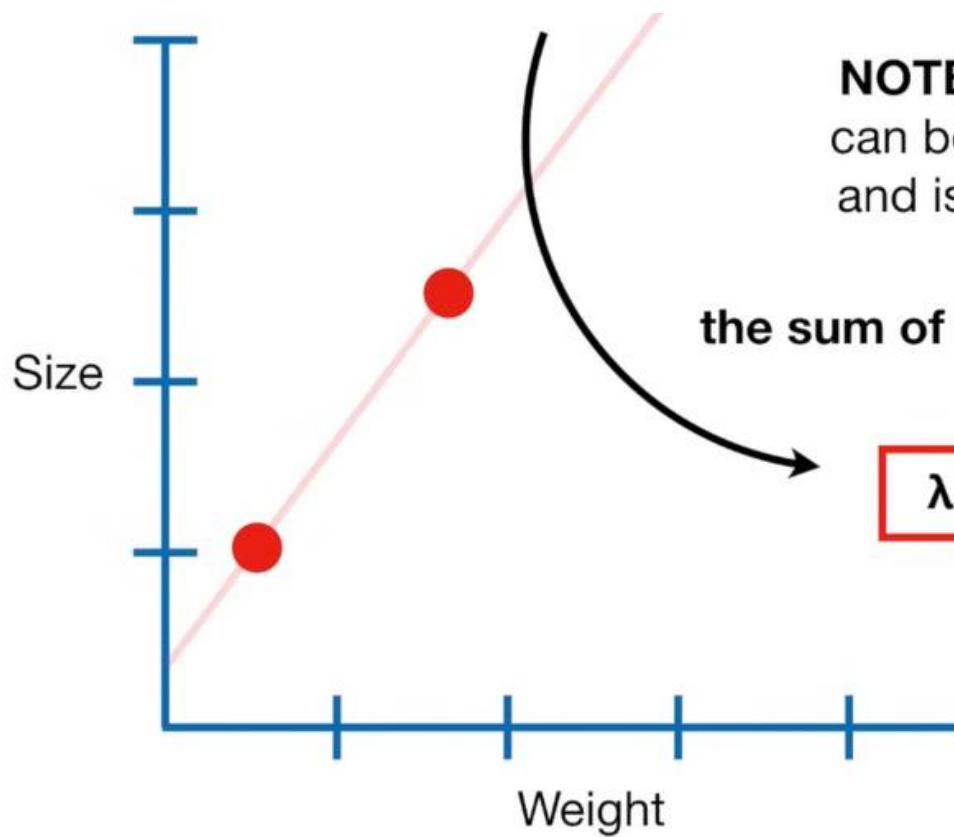
$\lambda \times (\text{slope1}^2 + \text{slope2}^2 + \text{slope3}^2 + \dots + \text{slope10000}^2)$

Lasso regression

Ridge Regression is just Least Squares...
...plus the Ridge Regression Penalty



Least **Absolute** Shrinkage and Selection Operator (LASSO)



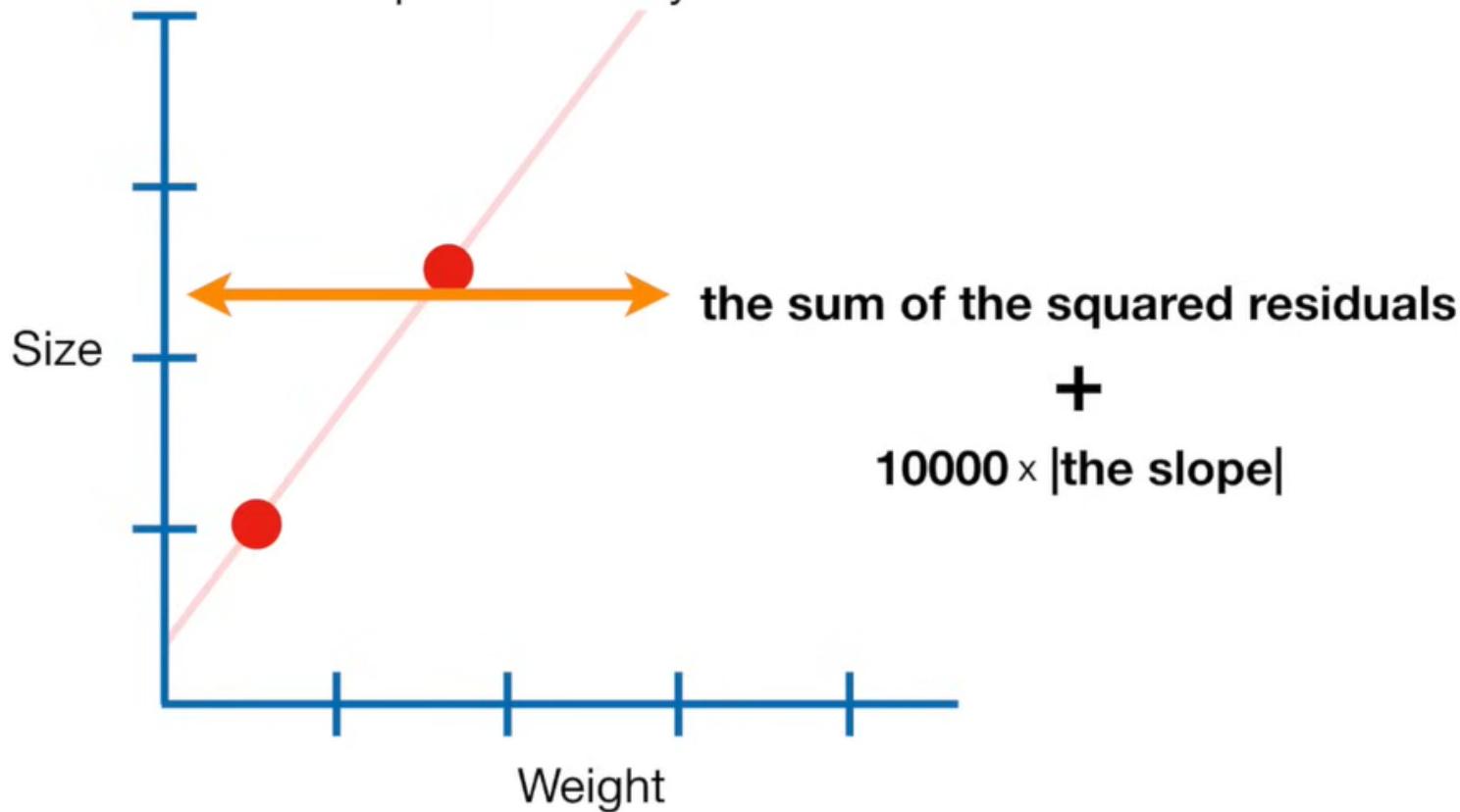
NOTE: Just like with **Ridge Regression**, λ can be any value from **0** to **positive infinity** and is determined using **Cross Validation**.

the sum of the squared residuals

+

$\lambda \times |\text{the slope}|$

The big difference between **Ridge** and **Lasso Regression** is that **Ridge Regression** can only shrink the slope asymptotically close to **0** while **Lasso Regression** can shrink the slope all the way to **0**.



Size = y-intercept + **slope** × **Weight** + **diet difference** × **High Fat Diet**
+ astrological offset × **Sign** + airspeed scalar × **Airspeed of Swallow**

...and the larger we make λ ...

...these parameters might shrink a little bit...

λ ($\text{slope}^2 + \text{diet difference}^2 + \text{astrological offset}^2 + \text{airspeed scalar}^2$)

Size = $y\text{-intercept} + \text{slope} \times \text{Weight} + \text{diet difference} \times \text{High Fat Diet}$

+ astrological offset \times **Sign** + airspeed scalar \times **Airspeed of Swallow**

...and the larger we make λ ...

...and these parameters might shrink a lot, but they will never be equal to 0.

λ

($\text{slope}^2 + \text{diet difference}^2 + \text{astrological offset}^2 + \text{airspeed scalar}^2$)

Size = y-intercept + slope × **Weight** + diet difference × **High Fat Diet**

- astrological offset × **Sign** + airspeed scalar × **Airspeed of Swallow**

...when we increase the
value for λ ...

...and these parameters will go all the
way to 0...

λ

($|\text{slope}| + |\text{diet difference}| + |\text{astrological offset}| + |\text{airspeed scalar}|$)



L.B.B

Laboratory of Systems Biology
and Bioinformatics

Ridge Regression is very similar to...

$$\begin{aligned} & \text{the sum of the squared residuals} \\ & + \\ & \lambda \times \text{the slope}^2 \end{aligned}$$

...Lasso Regression

$$\begin{aligned} & \text{the sum of the squared residuals} \\ & + \\ & \lambda \times |\text{the slope}| \end{aligned}$$

Size = y-intercept + slope × **Weight** + diet difference × **High Fat Diet**
~~+ astrological offset × **Sign** + airspeed scalar × **Airspeed of Swallow**~~

But the big difference is that **Lasso Regression** can exclude useless variables from equations.

Ridge :

$$\min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

Lasso :

$$\min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\},$$

ELASTIC NET

$$\min_{\beta \in \mathbb{R}^p} \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p \mathbf{X}_{ij} \beta_j)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}$$

the sum of the squared residuals

+

$\lambda_1 \times |\text{variable}_1| + \dots + |\text{variable}_x|$ + $\lambda_2 \times \text{variable}_1^2 + \dots + \text{variable}_x^2$

MACHINE LEARNING TECHNIQUES FOR SURVIVAL PREDICTION.

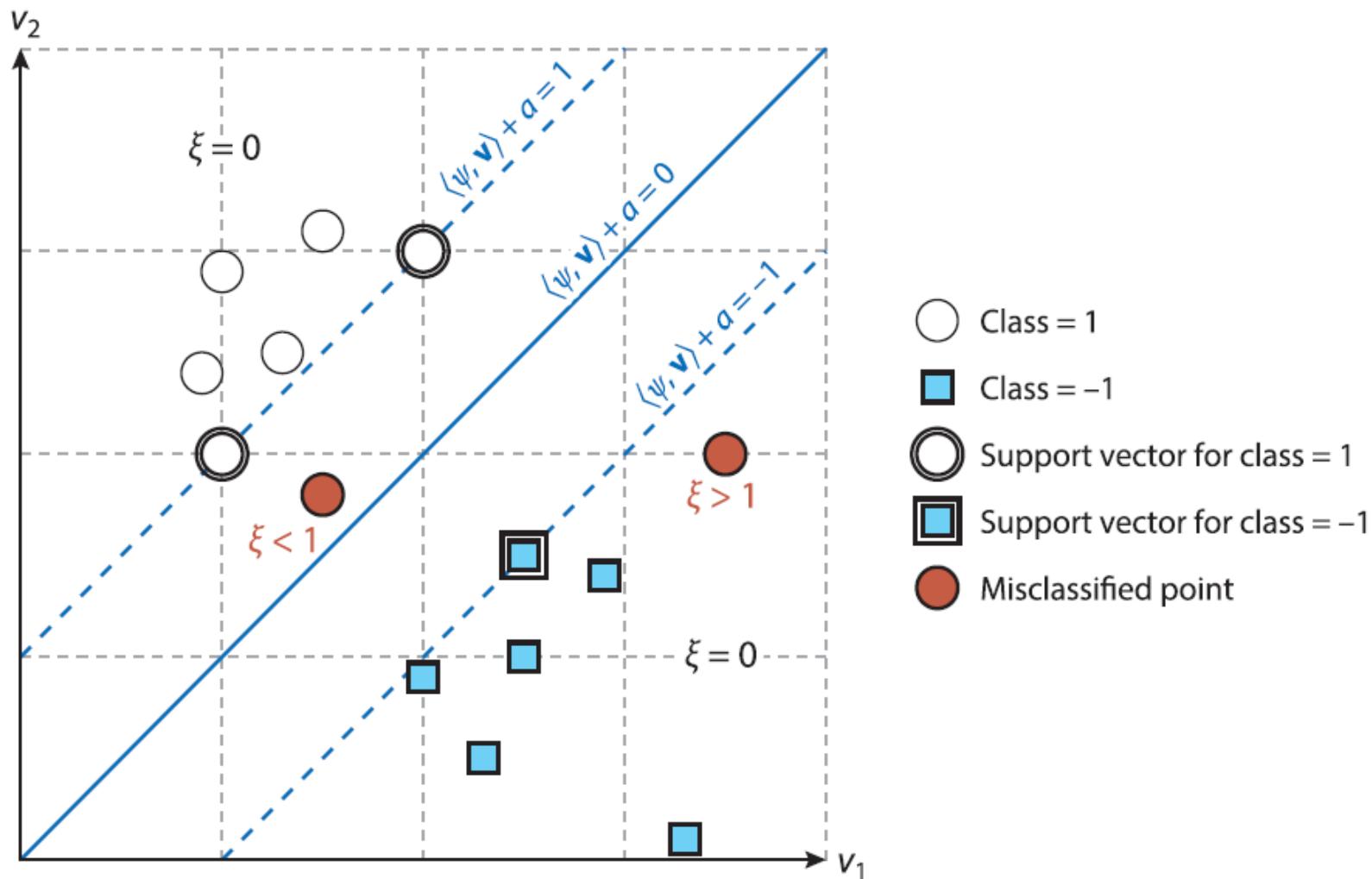
Support
Vector
Machines
(SVM)

TreeBased
Methods

Ensemble
Learners

Deep
Learning
Artificial
Neural
Networks

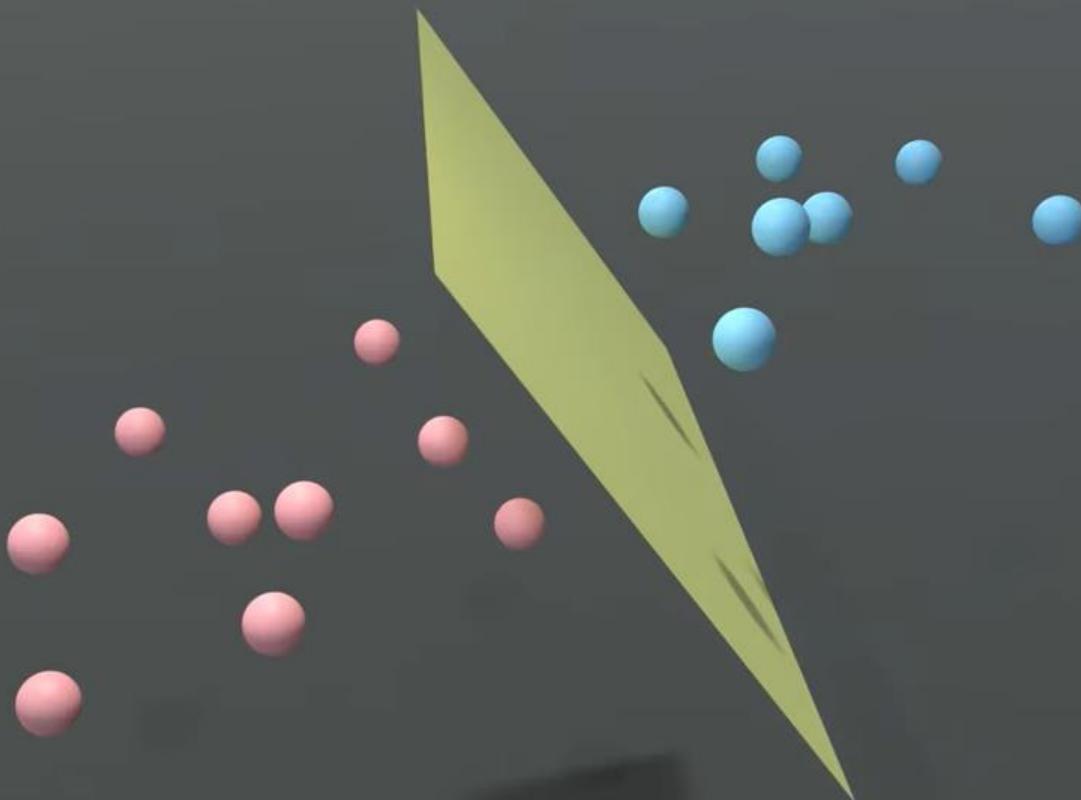
SVM





L.B.B

Laboratory of Systems Biology
and Bioinformatics



Point	Feature 1	Feature 2	...	Category
1	0.97	0.27		
2	0.77	0.19		
3	0.33	0.85		
4	0.41	0.93		
5	0.27	0.32		

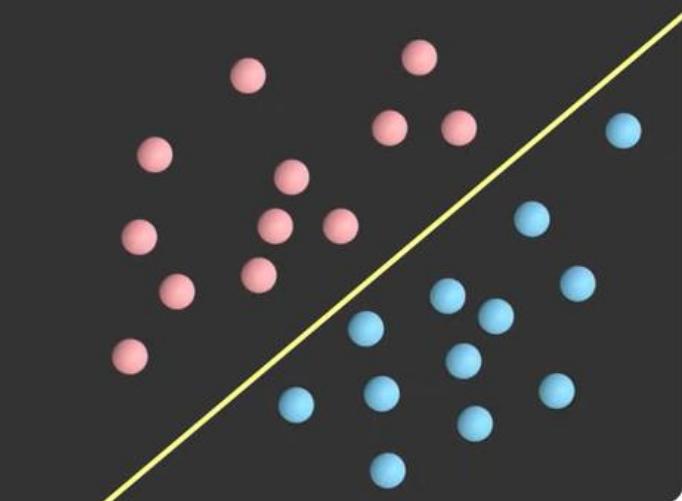
Supervised Learning Algorithm

Convex Optimization Problem

$$\max_{w,b} \|w\|^{-2}$$

$$w^T x + b \geq -1 \quad \forall x \in C_1$$

$$w^T x + b \leq 1 \quad \forall x \in C_2$$





L.B.B

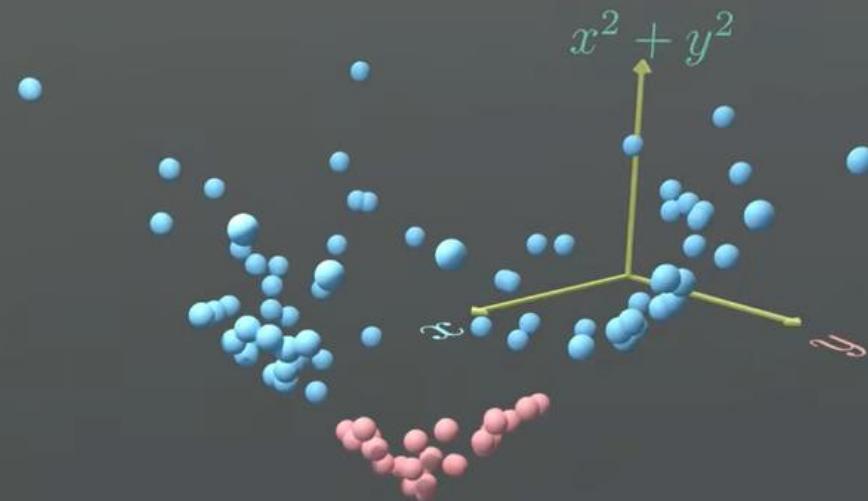
Laboratory of Systems Biology
and Bioinformatics





L.B.B

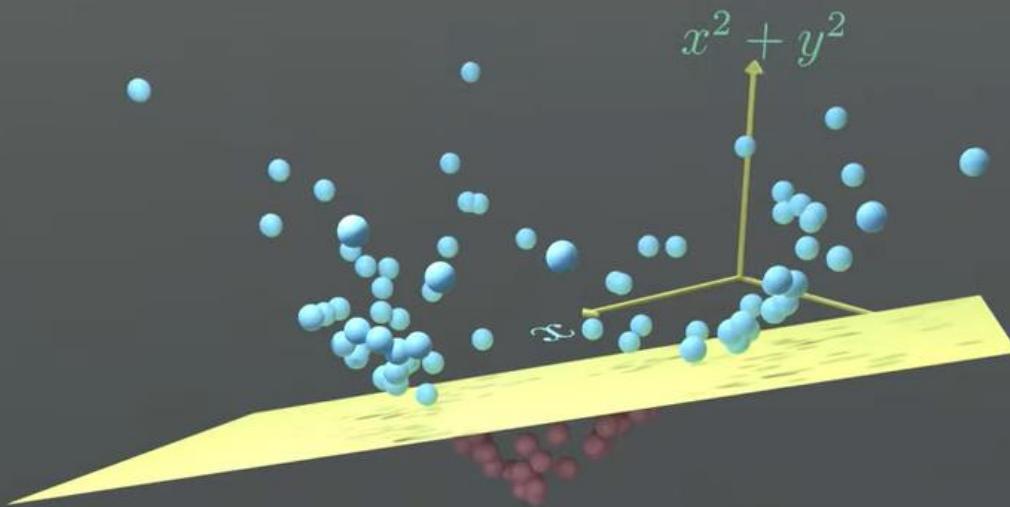
Laboratory of Systems Biology
and Bioinformatics





L.B.B

Laboratory of Systems Biology
and Bioinformatics



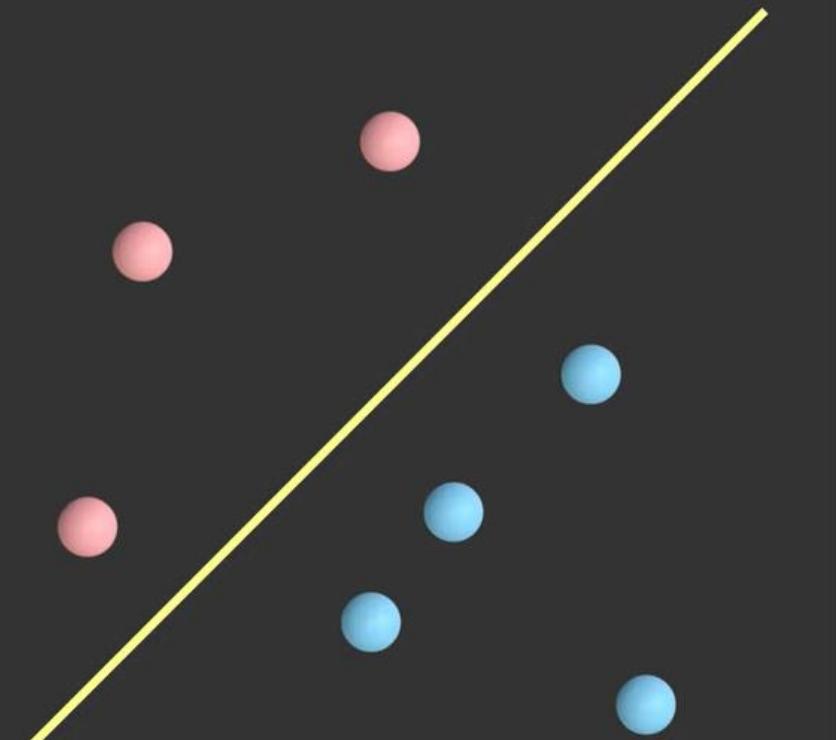


L.B.B

Laboratory of Systems Biology
and Bioinformatics

Kernel Trick

```
1 from sklearn import svm
2
3 # features
4 X = [
5     [-3, -1],
6     [0, -2],
7     [-2.5, 2],
8     [-1, -1],
9     [3, .5],
10    [.5, 3],
11    [-3, -3],
12 ]
13
14 # labels
15 y = [0, 1, 0, 1, 1, 0, 1]
16
17
18 # fit
19 clf = svm.SVC(kernel='linear').fit(X, y)
20
21
22 # predict
23 clf.predict([[2, 4]]) => 0
```



MACHINE LEARNING TECHNIQUES FOR SURVIVAL PREDICTION.

Support
Vector
Machines
(SVM)

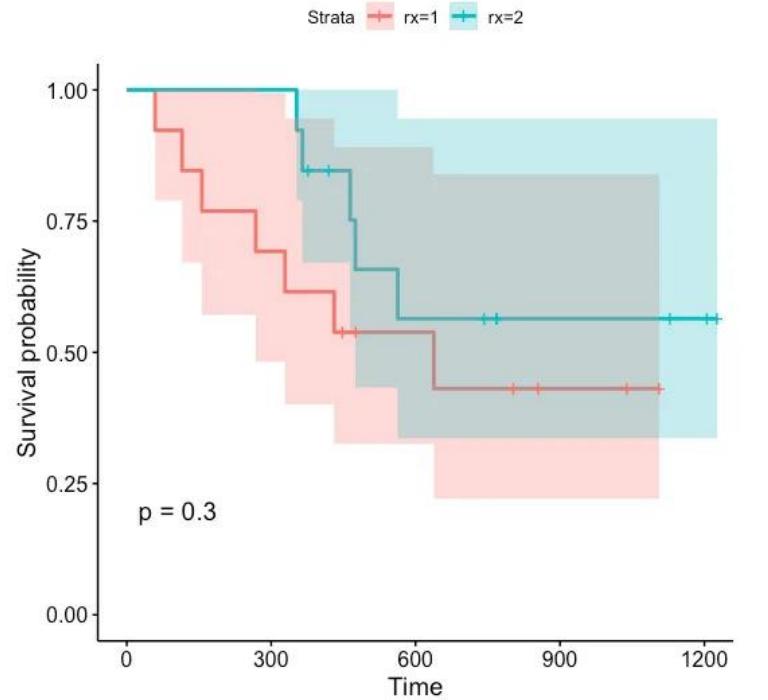
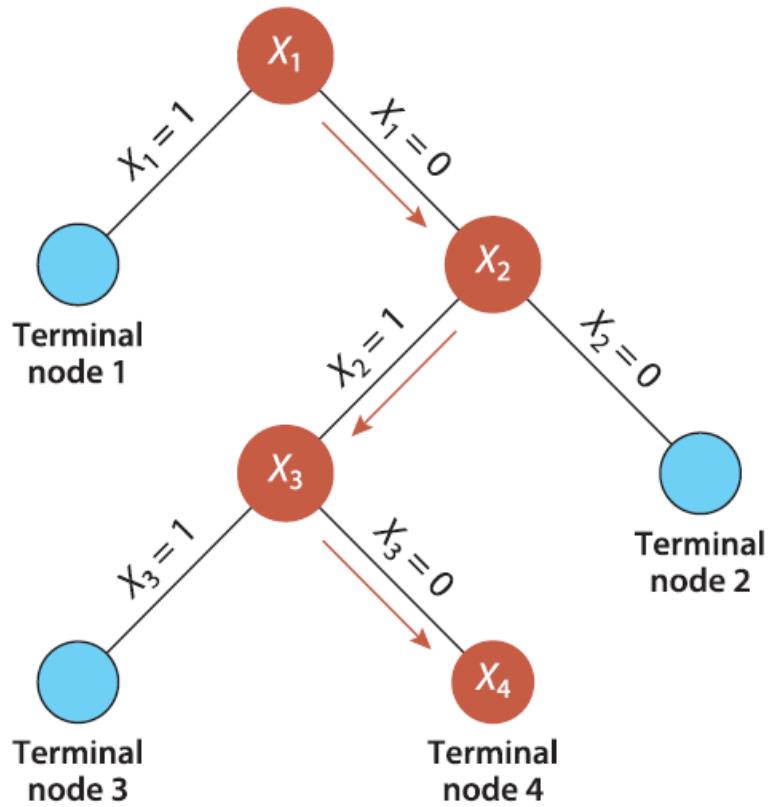
**TreeBased
Methods**

Ensemble
Learners

Deep
Learning
Artificial
Neural
Networks

Tree based methods

- Gordon & Olshen (1985)
- Ciampi et al. (1986, 1987)



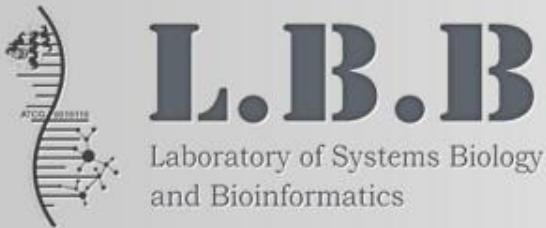
```

library(survival)
library(survminer)
data(ovarian)fit <- survfit(Surv(ovarian$futime,
ovarian$fustat) ~ rx,
data = ovarian, conf.type="log")\l
p_fit <- ggsurvplot(
fit, conf.int = TRUE, pval = TRUE)
p_fit
  
```

Algorithm

1. Discretize each covariate to be a *binary* variable.
2. For every binary covariate, X_j , $j = 1, \dots, p$, compute the log-rank statistic to test the difference between the survival curves for the two groups defined by X_j .
3. Choose the covariate—say, X_j —with the largest significant test statistic and partition the full sample into two groups (*child nodes*) based on X_j .
4. Repeat steps 2 and 3 for each subset (child node) until reaching the *terminal nodes*

That is, no covariates produce a significant test statistic and there are enough events in each *terminal node*.



Laboratory of Systems Biology
and Bioinformatics

Corpus ID: 61172913

Tree-based methods for survival analysis and high-dimensional data

Ruoqing Zhu • Published 1 December 2013 • Computer Science, Mathematics

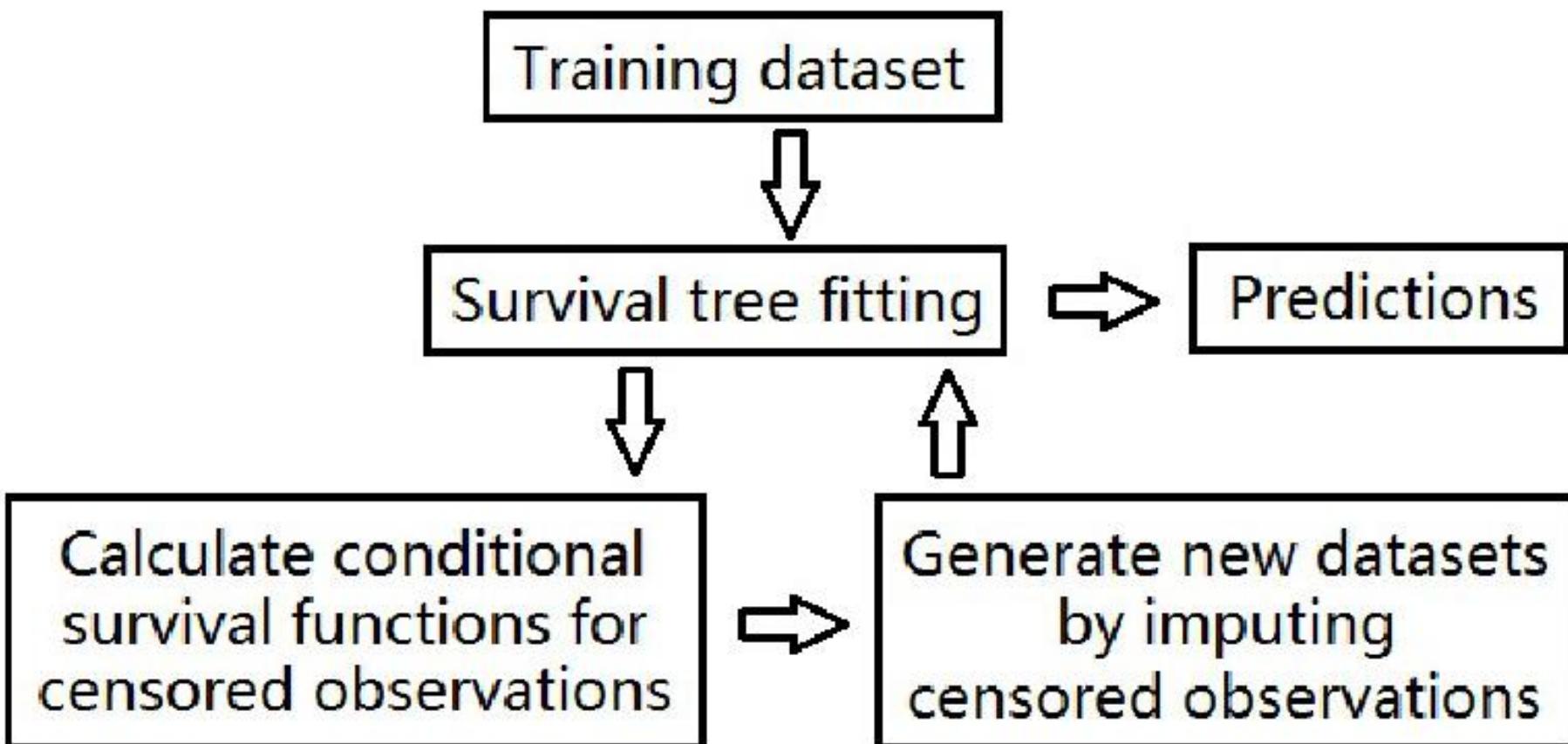
TLDR A central idea of this dissertation is focused on increasing the chance of using signaled variables as splitting rule during the tree construction while not loosing the randomness/diversity, hence a more accurate model can be built.[Expand](#)

cdr.lib.unc.edu

[Save to Library](#)

[Create Alert](#)

[Cite](#)



MACHINE LEARNING TECHNIQUES FOR SURVIVAL PREDICTION.

Support
Vector
Machines
(SVM)

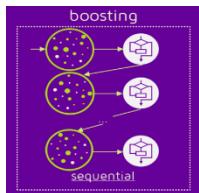
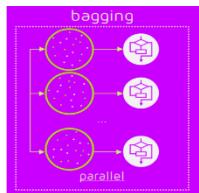
TreeBased
Methods

**Ensemble
Learners**

Deep
Learning
Artificial
Neural
Networks

Ensemble learner

- Bagging.
- Boosting.
- Random forests.



BAGGING

Machine Learning, 24, 123–140 (1996)

© 1996 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Bagging Predictors

LEO BREIMAN

Statistics Department, University of California, Berkeley, CA 94720

leo@stat.berkeley.edu

Editor: Ross Quinlan

Abstract. Bagging predictors is a method for generating multiple versions of a predictor and using these to get an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome and does a plurality vote when predicting a class. The multiple versions are formed by making bootstrap replicates of the learning set and using these as new learning sets. Tests on real and simulated data sets using classification and regression trees and subset selection in linear regression show that bagging can give substantial gains in accuracy. The vital element is the instability of the prediction method. If perturbing the learning set can cause significant changes in the predictor constructed, then bagging can improve accuracy.

Keywords: Aggregation, Bootstrap, Averaging, Combining

SPRINGER LINK

[Find a journal](#)[Publish with us](#)[Search](#)

[Home](#) > [Machine Learning](#) > Article

[Published: August 1996](#)

Bagging predictors

[Leo Breiman](#)

[Machine Learning](#) **24**, 123–140 (1996) | [Cite this article](#)

67k Accesses | **12k** Citations | **44** Altmetric | [Metrics](#)

Comment > Stat Med. 2004 Jan 15;23(1):77-91. doi: 10.1002/sim.1593.

Bagging survival trees

Torsten Hothorn ¹, Berthold Lausen, Axel Benner, Martin Radespiel-Tröger

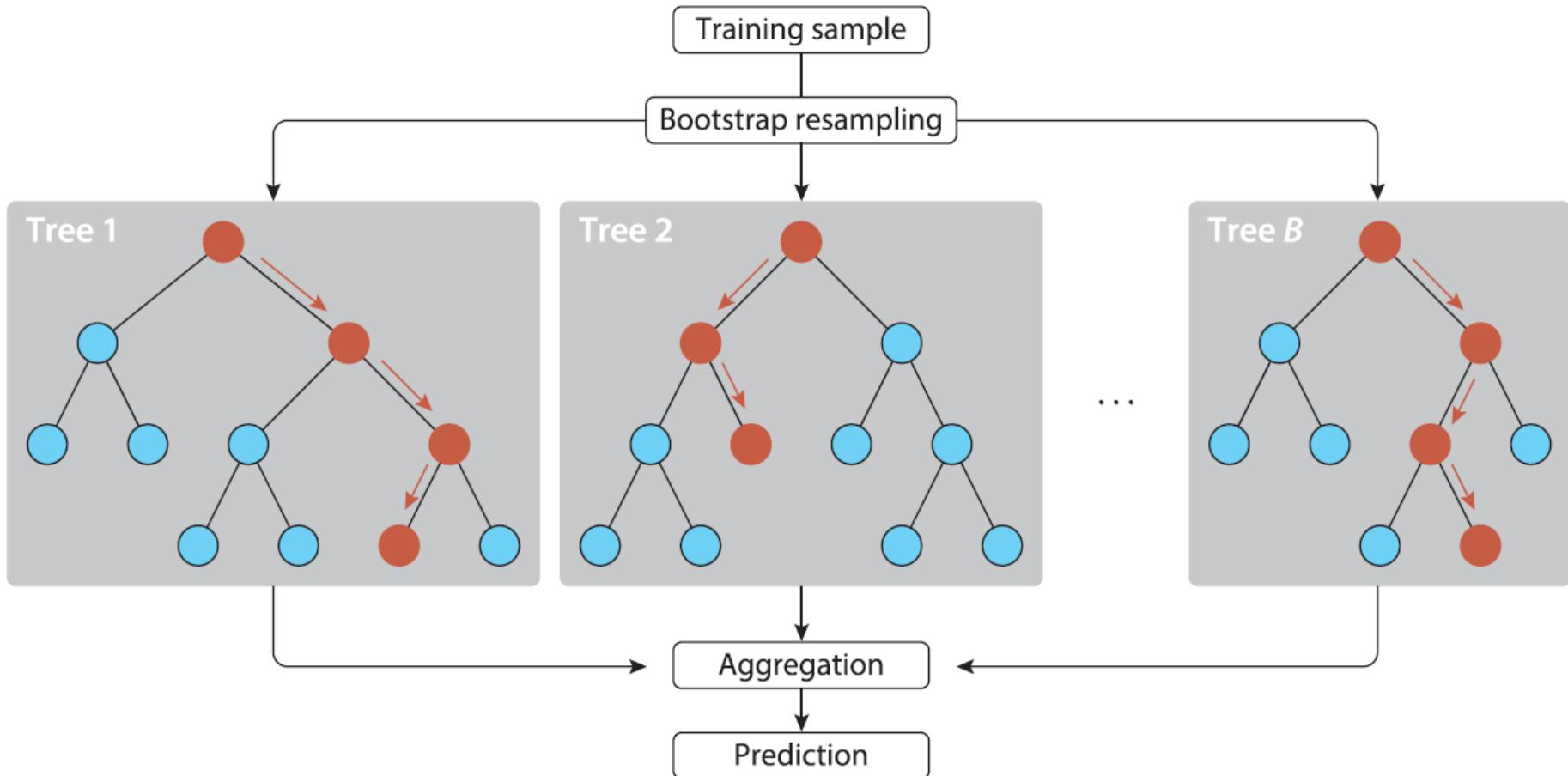
Affiliations + expand

PMID: 14695641 DOI: 10.1002/sim.1593



L.B.B

Laboratory of Systems Biology
and Bioinformatics



- Training weak learners (e.g., individual survival trees) in parallel, and combining Results .
- Majority voting

Boosting

Statistical Science

2007, Vol. 22, No. 4, 477–505

DOI: 10.1214/07-STS242

© Institute of Mathematical Statistics, 2007

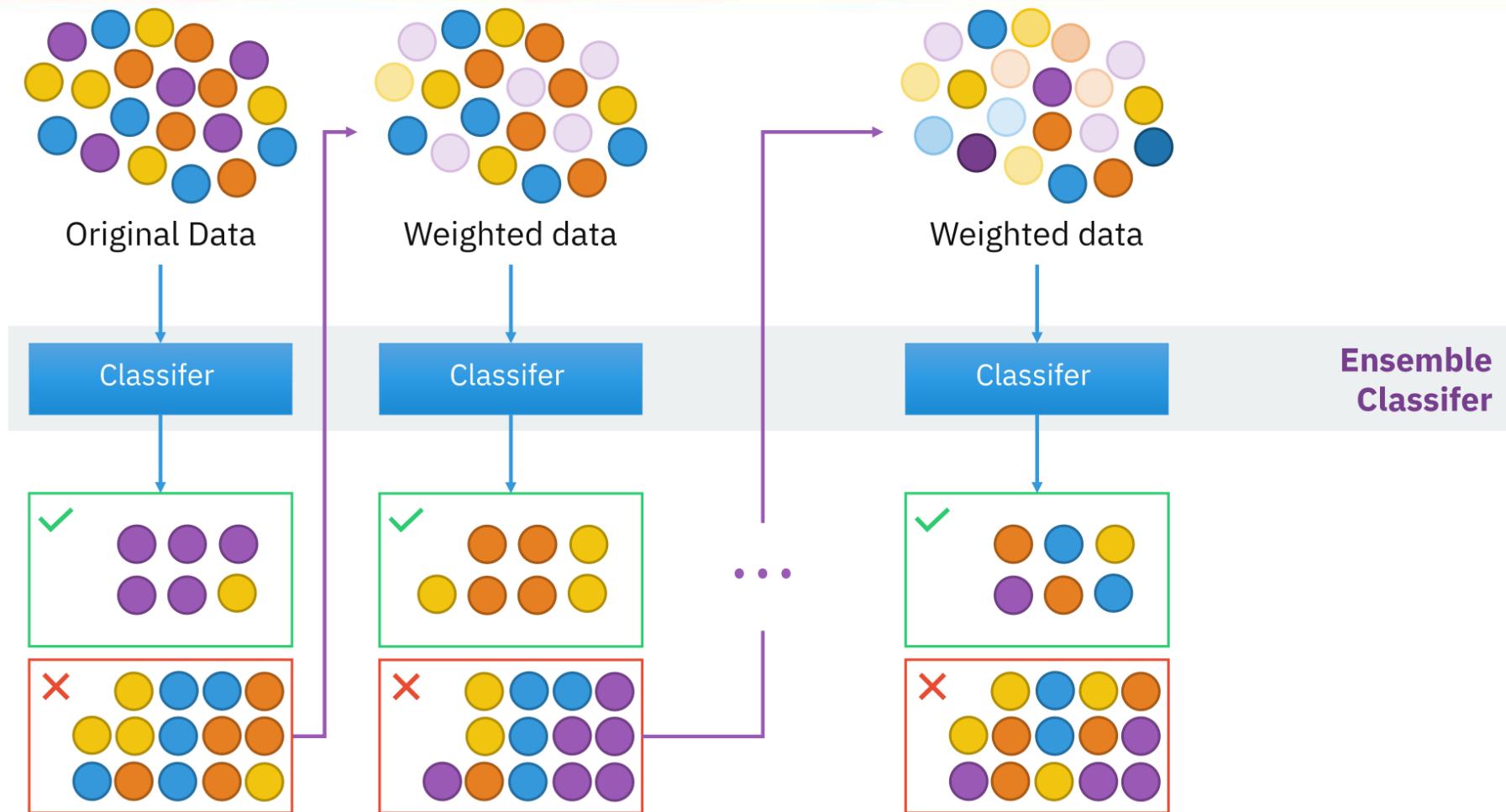
Boosting Algorithms: Regularization, Prediction and Model Fitting

Peter Bühlmann and Torsten Hothorn



L.B.B

Laboratory of Systems Biology
and Bioinformatics



$$\mathcal{F}_m(\mathbf{X}_i) = \mathcal{F}_{m-1}(\mathbf{X}_i) + w_m f_m(\mathbf{X}_i),$$

Random forests.



Machine Learning, 45, 5–32, 2001
© 2001 Kluwer Academic Publishers. Manufactured in The Netherlands.

Random Forests

LEO BREIMAN

Statistics Department, University of California, Berkeley, CA 94720

Editor: Robert E. Schapire

<https://blog.faradars.org/random-forest-algorithm/>

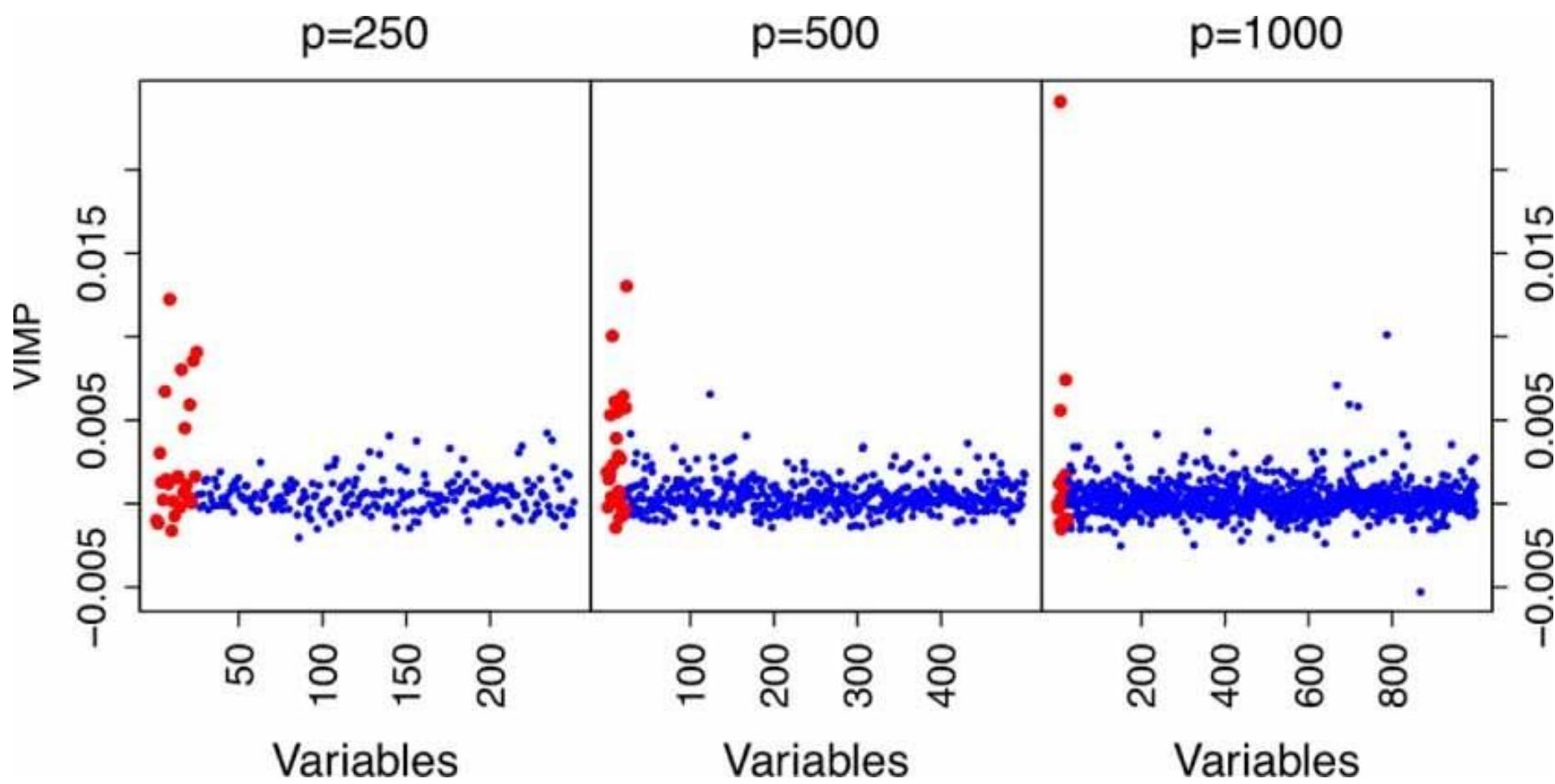
STATISTICAL ANALYSIS AND DATA MINING

Research Article

Random survival forests for high-dimensional data

Hemant Ishwaran , Udaya B. Kogalur, Xi Chen, Andy J. Minn

First published: 11 January 2011 | <https://doi.org/10.1002/sam.10103> | Citations: 118



DEEP LEARNING
AND
ARTIFICIAL NEURAL NETWORKS

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR
INFORMATION STORAGE AND ORGANIZATION
IN THE BRAIN¹

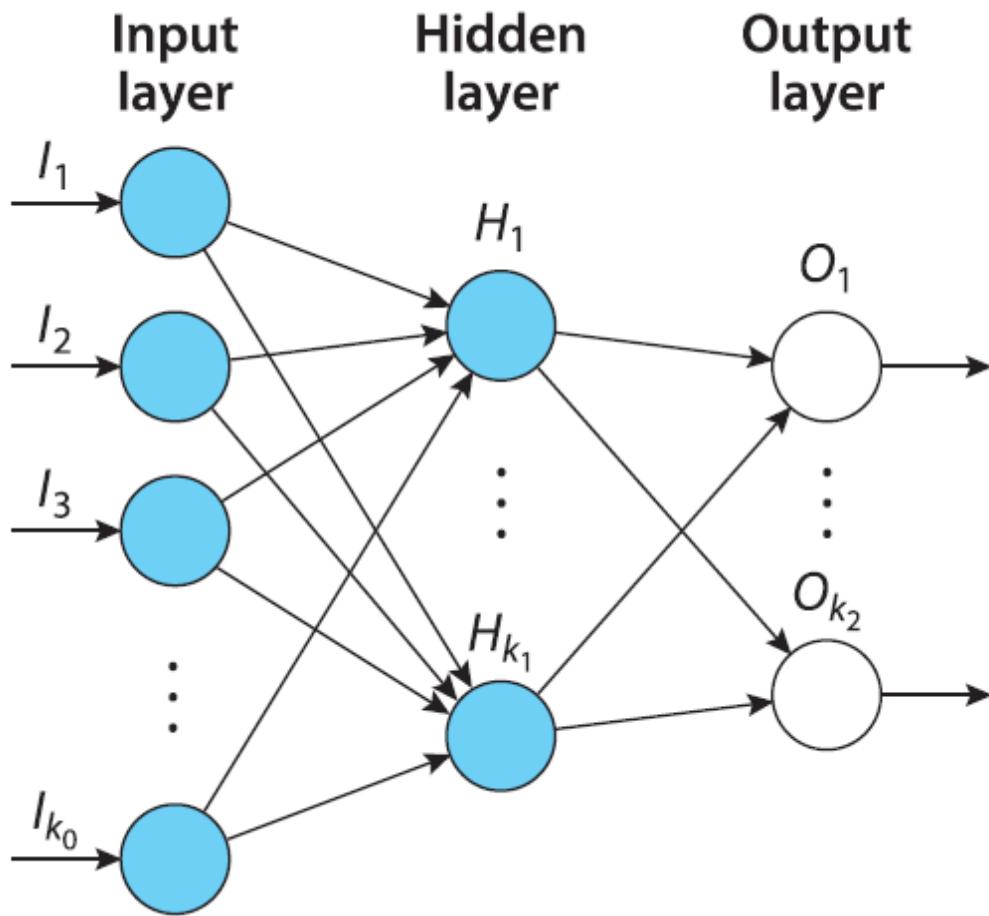
F. ROSENBLATT

Cornell Aeronautical Laboratory



L.B.B

Laboratory of Systems Biology
and Bioinformatics



Comparative Study

> Stat Med. 1995 Jan 15;14(1):73-82. doi: 10.1002/sim.4780140108.

A neural network model for survival data

D Faraggi¹, R Simon

Affiliations + expand

PMID: 7701159 DOI: [10.1002/sim.4780140108](https://doi.org/10.1002/sim.4780140108)

- Fully connected, feed-forward neural network .
- Survival status as a training label and output predicted survival probabilities.
- Bayesian networks
- Convolutional neural networks
- Recurrent neural networks (Yang et al. 2018).

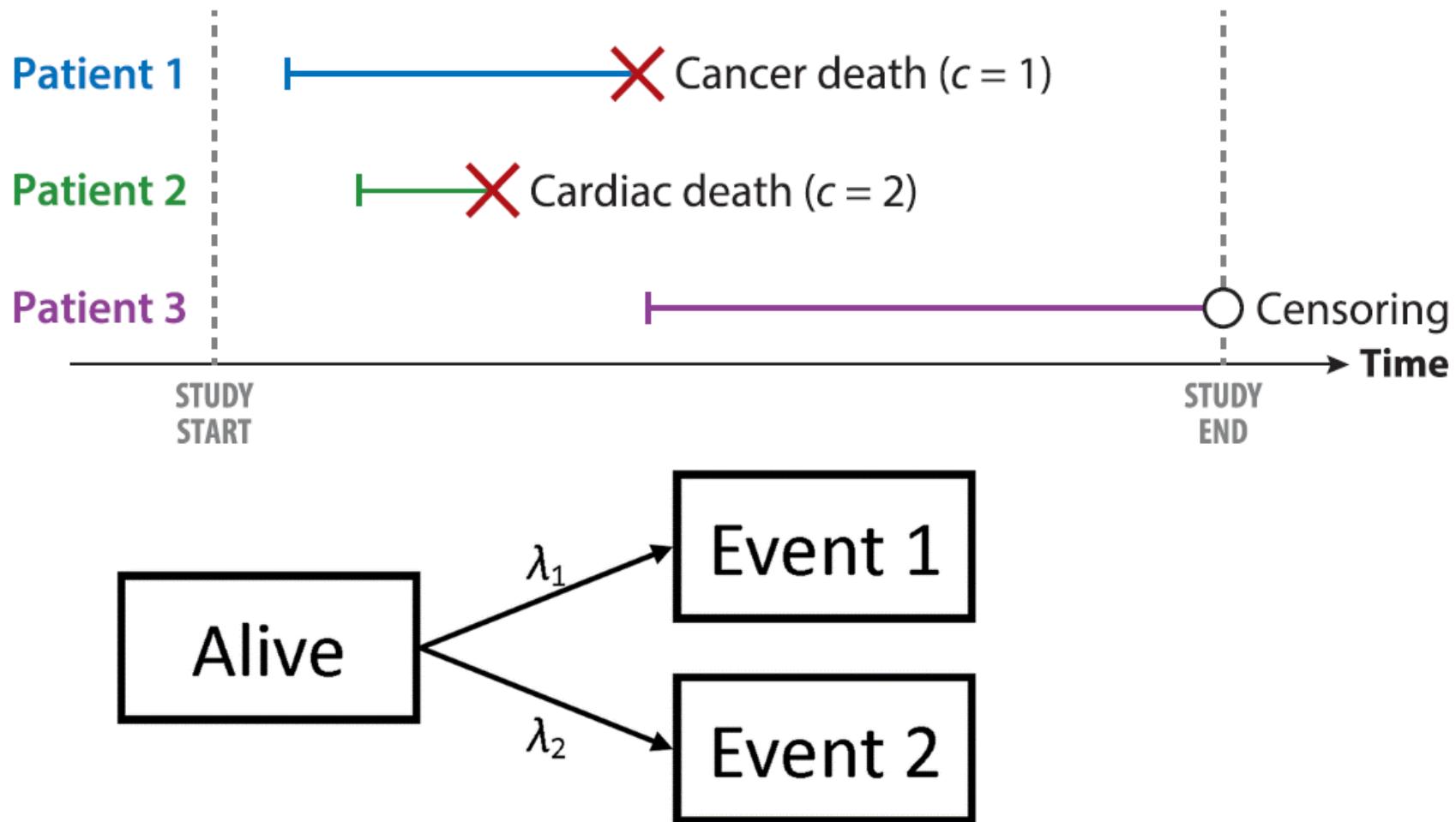
- PREDICTION FOR COMPETING AND SEMICOMPETING RISKS

Competing Risks

Semicompeting Risks



COMPETING RISK



The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks

Changhee Lee,¹ William R. Zame,² Jinsung Yoon,¹ Mihaela van der Schaar^{3, 1, 4}

¹ Department of Electrical and Computer Engineering, University of California, Los Angeles, USA

² Department of Economics, University of California, Los Angeles, USA

³ Department of Engineering Science, University of Oxford, UK

⁴ Alan Turing Institute, London, UK

chl8856@ucla.edu, zame@econ.ucla.edu, jsyoon0823@ucla.edu, mihaela.vanderschaar@oxford-man.ox.ac.uk

Package ‘cmprsk’

October 12, 2022

Version 2.2-11

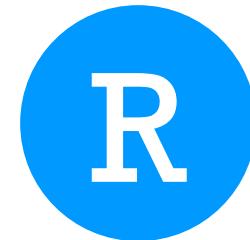
Date 2021-12-23

Title Subdistribution Analysis of Competing Risks

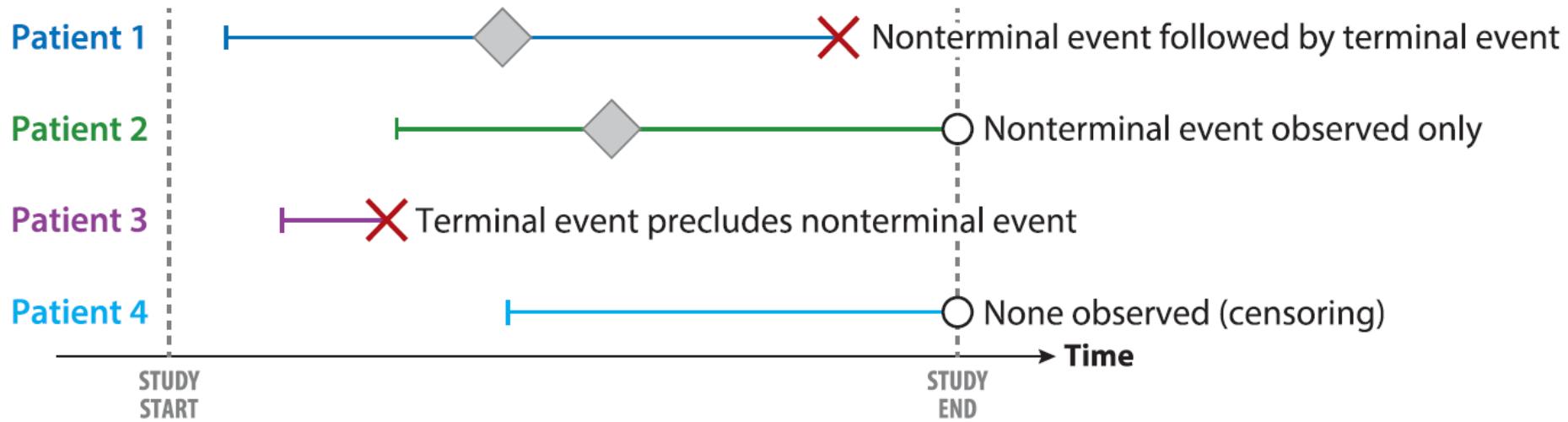
Author Bob Gray <gray@jimmy.harvard.edu>

Maintainer Bob Gray <gray@jimmy.harvard.edu>

Depends R (>= 3.0.0), survival



SEMICOMPETING RISKS



- Occurrence of one event, a **Nonterminal** event, is subject to the occurrence of another, **Terminal**.

As the Nonterminal event (e.g., **Cancer progression**) is often a strong precursor to the terminal event (**Death**),

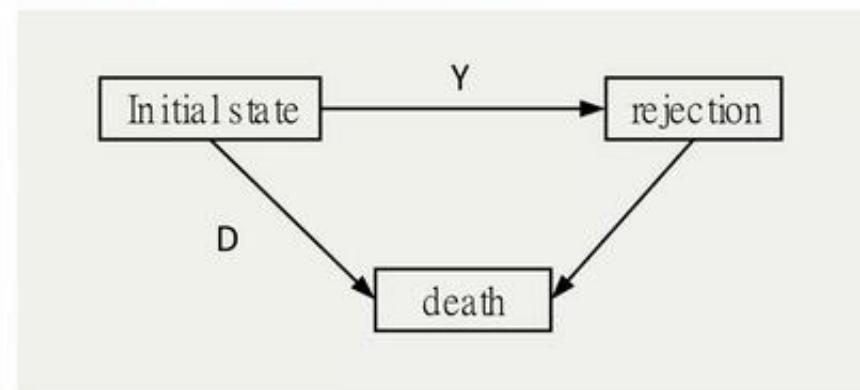
- Observed variables:

$$Y \wedge D \wedge C$$

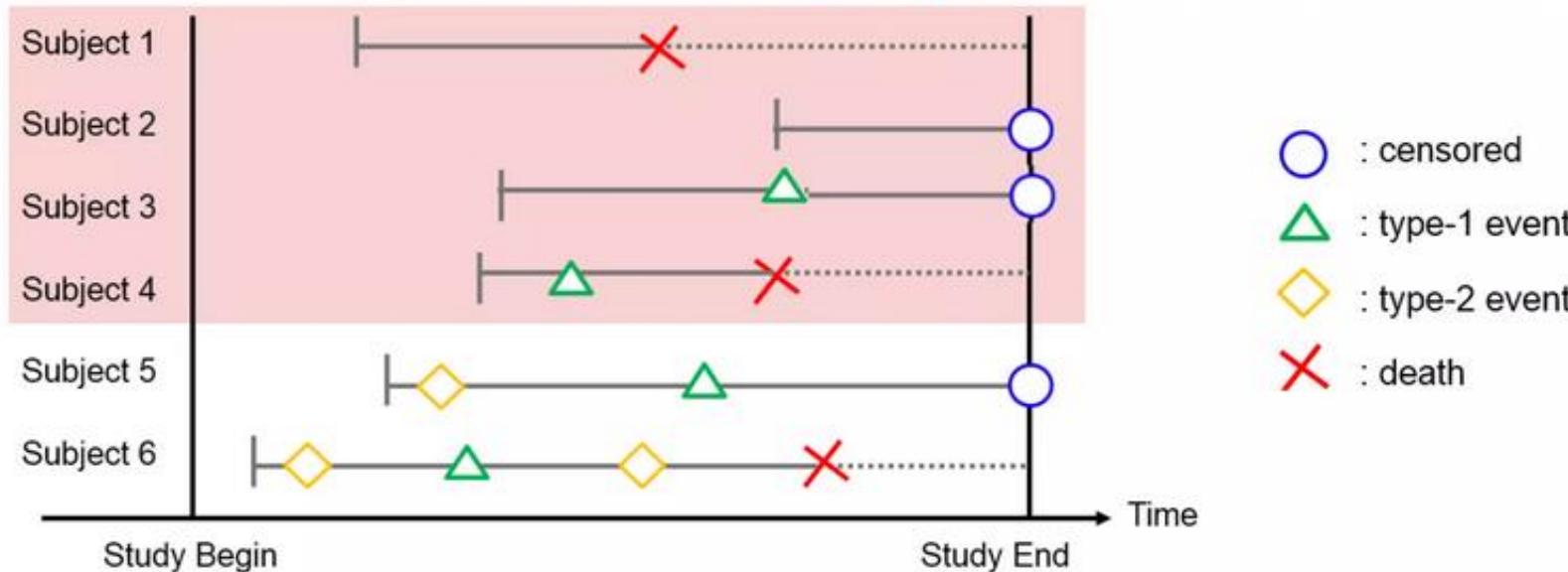
$$I(Y \leq D \wedge C)$$

$$D \wedge C$$

$$I(D \leq C)$$



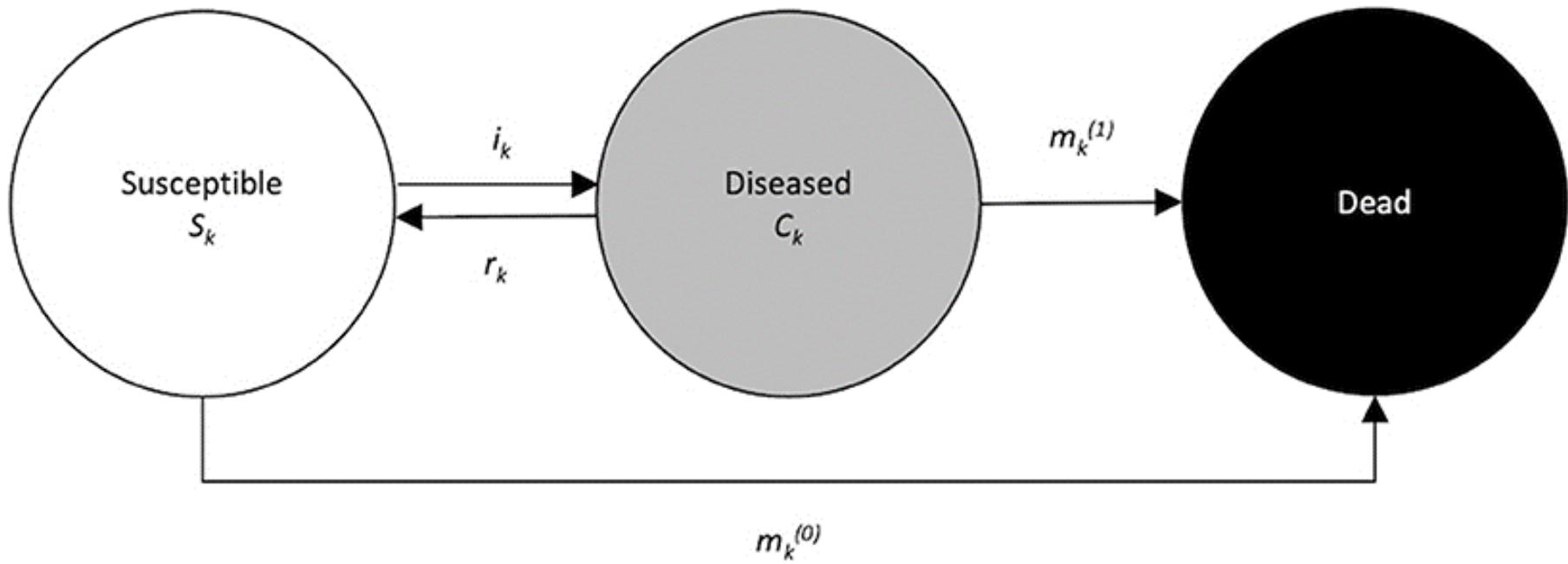
- Semi-competing risks: *D* is a competing risks for *Y* but not vice versa



What is lacking ?

How to model a predictor's potentially different roles in disease progress and death ?

The illness-death model.



Which individuals transition from being **event-free** (e.g., from time of diagnosis) to **progression** or to **death** or from **progression to death** .

The hazards for each subject at t :

$$\lambda_1(t | \gamma_i, X_i) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t \leq \overline{T}_{i1} < t + \Delta | T_{i1} \geq t, \overline{T}_{i2} \geq t, \gamma_i, X_i)}{\Delta}$$
$$= \gamma_i \lambda_{01}(t) \exp \{h_1(X_i)\}$$

Times to the nonterminal and terminal events and to the censoring time.



$$\lambda_1(t \mid \gamma_i, \mathbf{X}_i) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t \leq T_{i1} < t + \Delta \mid T_{i1} \geq t, T_{i2} \geq t, \gamma_i, \mathbf{X}_i)}{\Delta} = \gamma_i \lambda_{01}(t) \exp \{h_1(\mathbf{X}_i)\}, \quad 8.$$

$$\lambda_2(t \mid \gamma_i, \mathbf{X}_i) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t \leq T_{i2} < t + \Delta \mid T_{i1} \geq t, T_{i2} \geq t, \gamma_i, \mathbf{X}_i)}{\Delta} = \gamma_i \lambda_{02}(t) \exp \{h_2(\mathbf{X}_i)\}, \quad 9.$$

$$\lambda_3(t \mid t_1, \gamma_i, \mathbf{X}_i) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t \leq T_{i2} < t + \Delta \mid T_{i1} = t_1, T_{i2} \geq t, \gamma_i, \mathbf{X}_i)}{\Delta} = \gamma_i \lambda_{03}(t - t_1) \exp \{h_3(\mathbf{X}_i)\}, \quad 10.$$

The transitions from diagnosis to progression prior to death, (8)

From diagnosis to death prior to progression. (9)

from progression (that happens at t1) to death . (10)

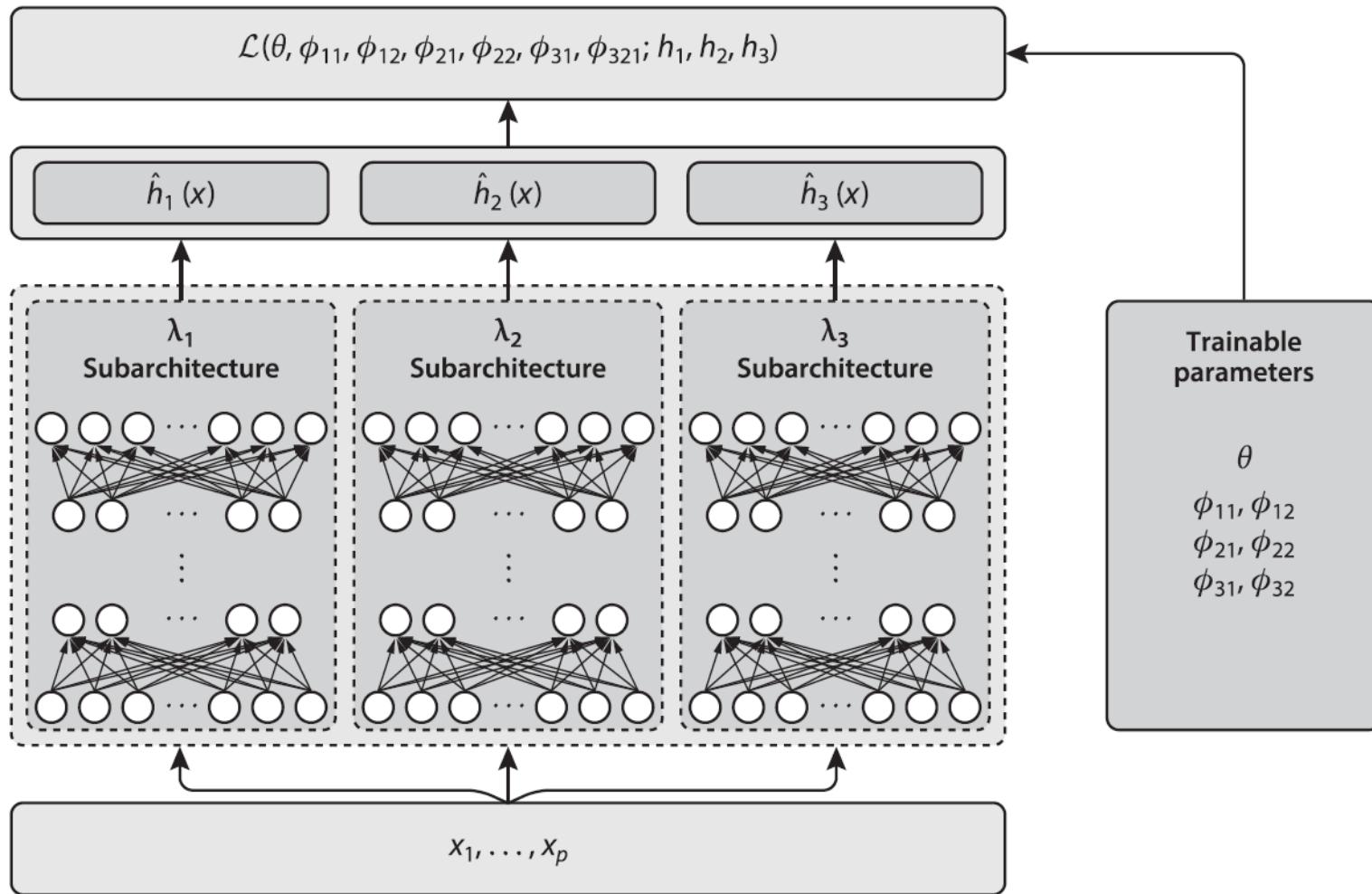


L.B.B

Laboratory of Systems Biology
and Bioinformatics

DNN-SCR

- A new deep learning approach for semi competing risks.

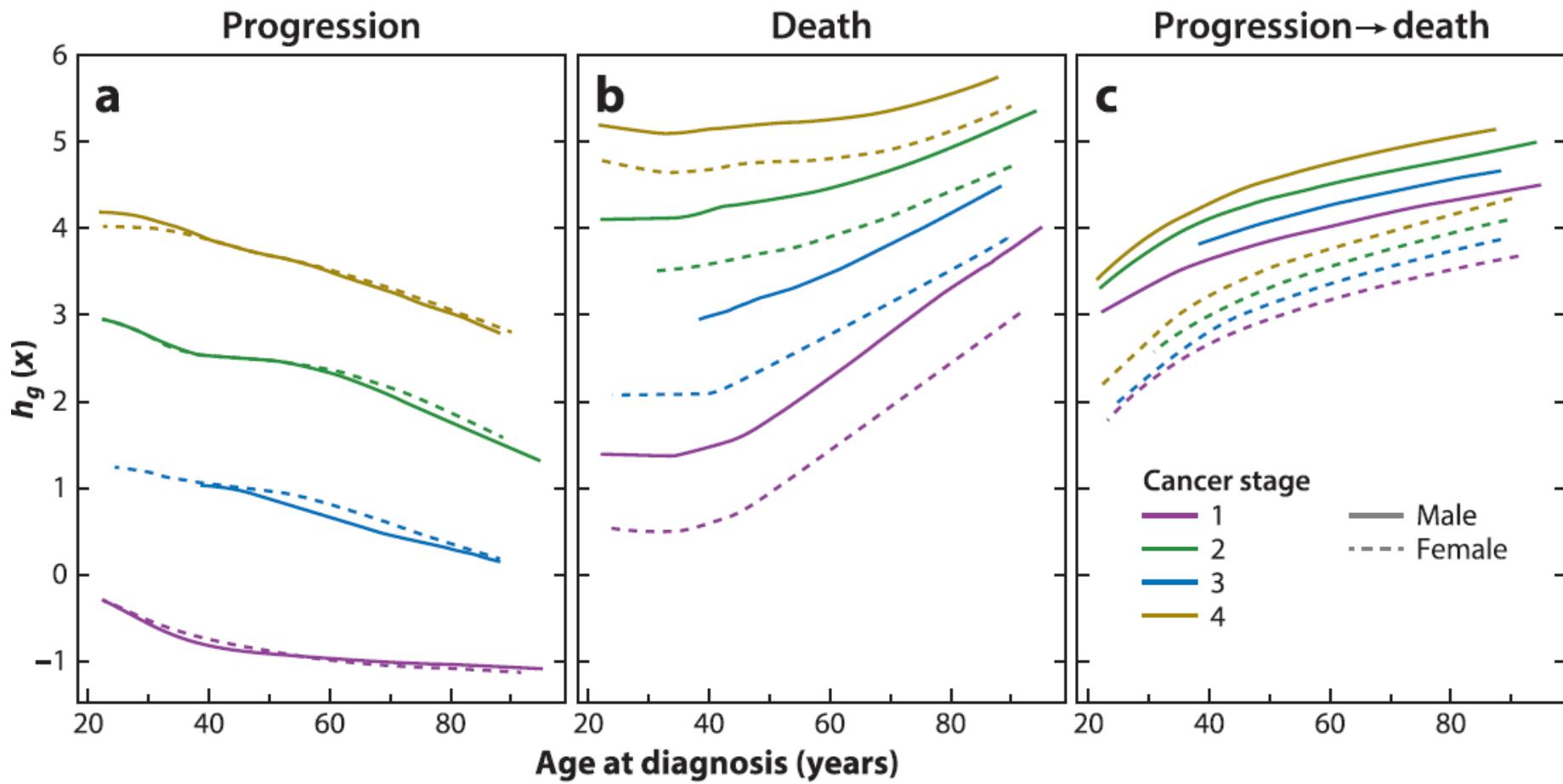


The subset includes **5,296** patients with **nonsmall cell lung cancer** diagnosed between June 1983 and October 2021.

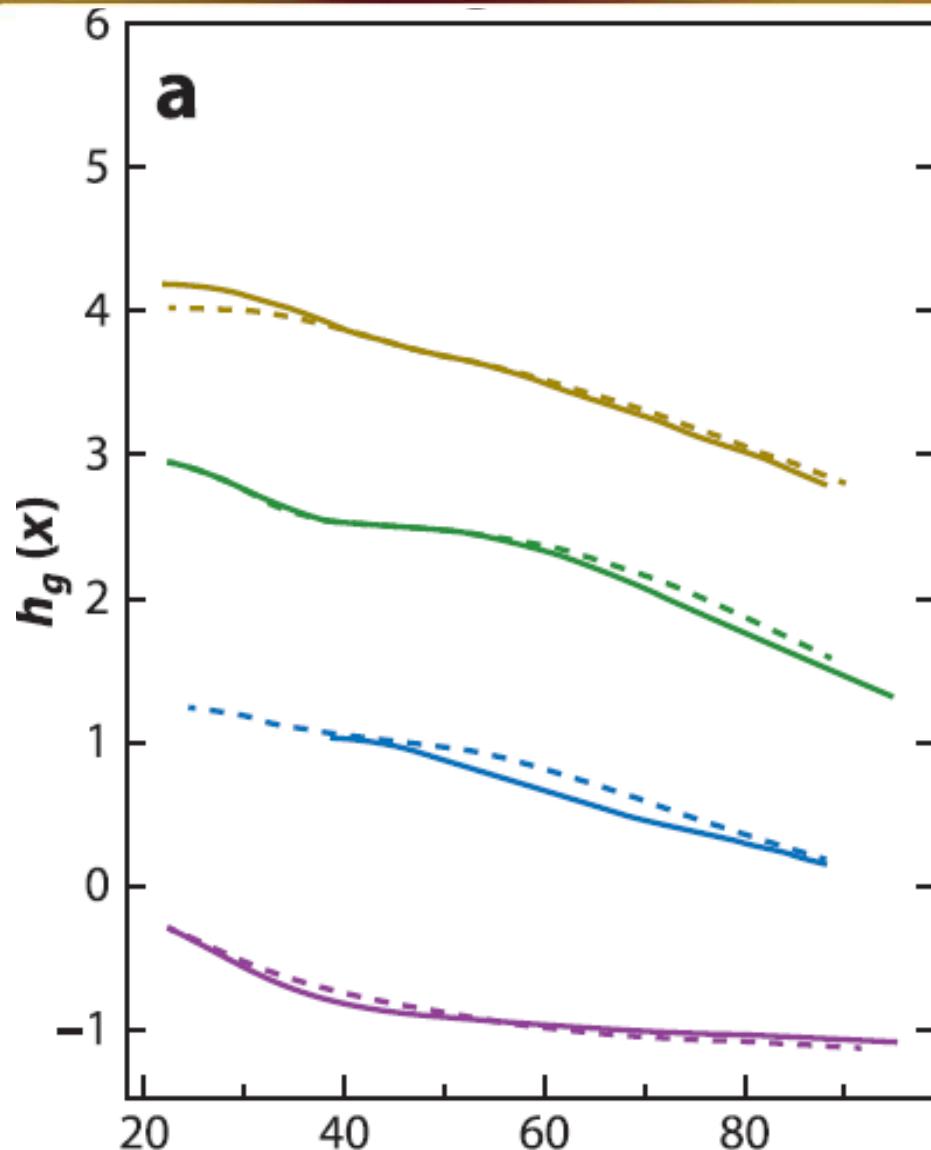
- Age at diagnosis (years),
- Sex (0: male, 1: female),
- Race (0: other, 1: white),
- Ethnicity (0: non-hispanic, 1: hispanic),
- Height (meters),
- Weight (kilograms),
- Smoking status (0: never, 1: former, 2: current),pack-years,
- Cancer stage (1–4),
- Two indicators of genetic mutations [epidermal growth factor receptor (EGFR) and kirsten rat sarcoma viral oncogene homolog (KRAS)]

SEMICOMPETING

- The date of progression is the date of the first source evidence, including exam, Radiology report or pathology.
- *Progression* but alive is observed in 224 patients (4%)
- *Death* prior to progression is observed among 1,916 patients (36%).
- *Progression* followed by *death* is observed in 111 patients (2%),



Progression

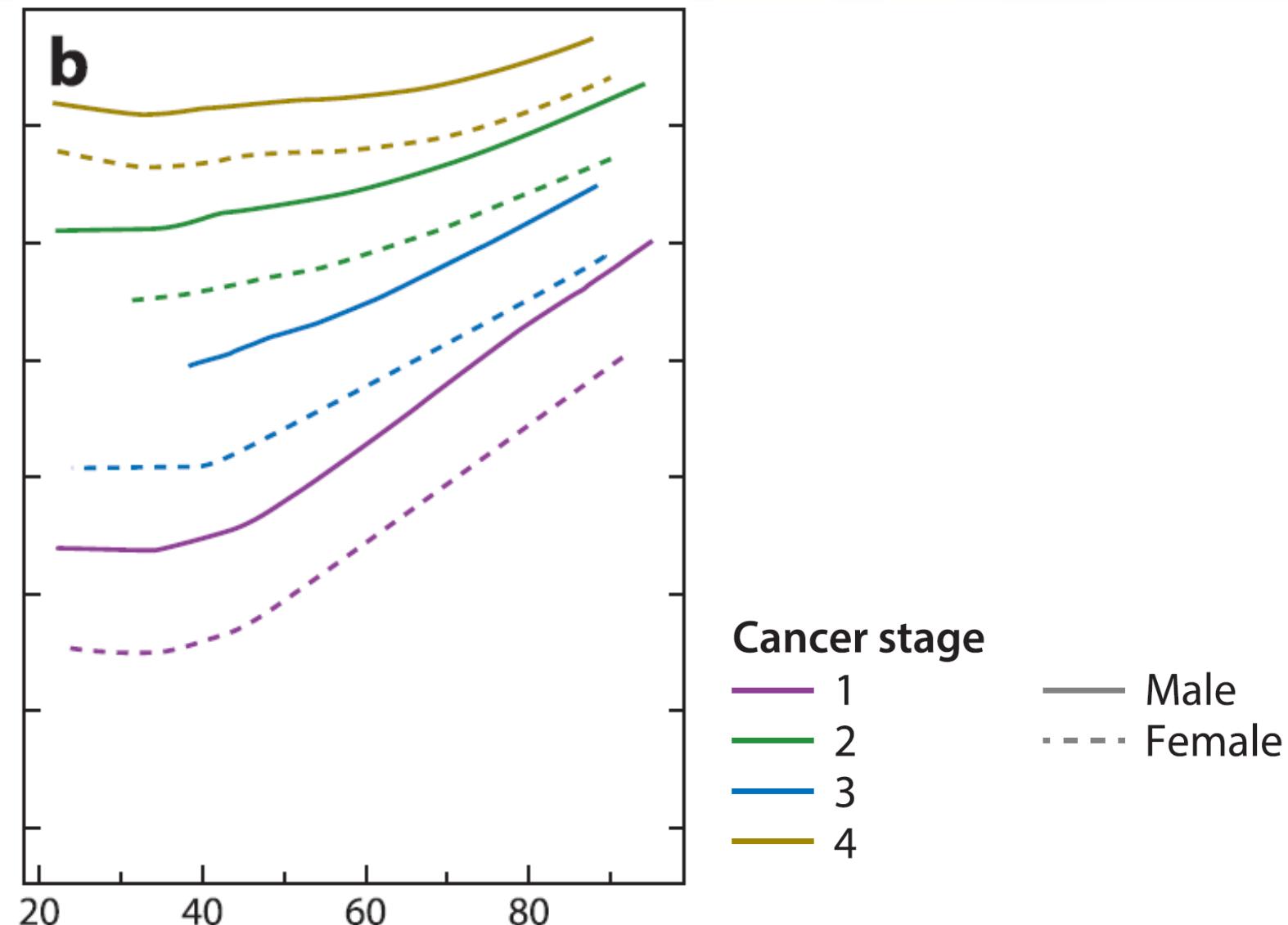


shows that younger age and more advanced stage are associated with higher hazards for progression; for the transitions from diagnosis or progression to death

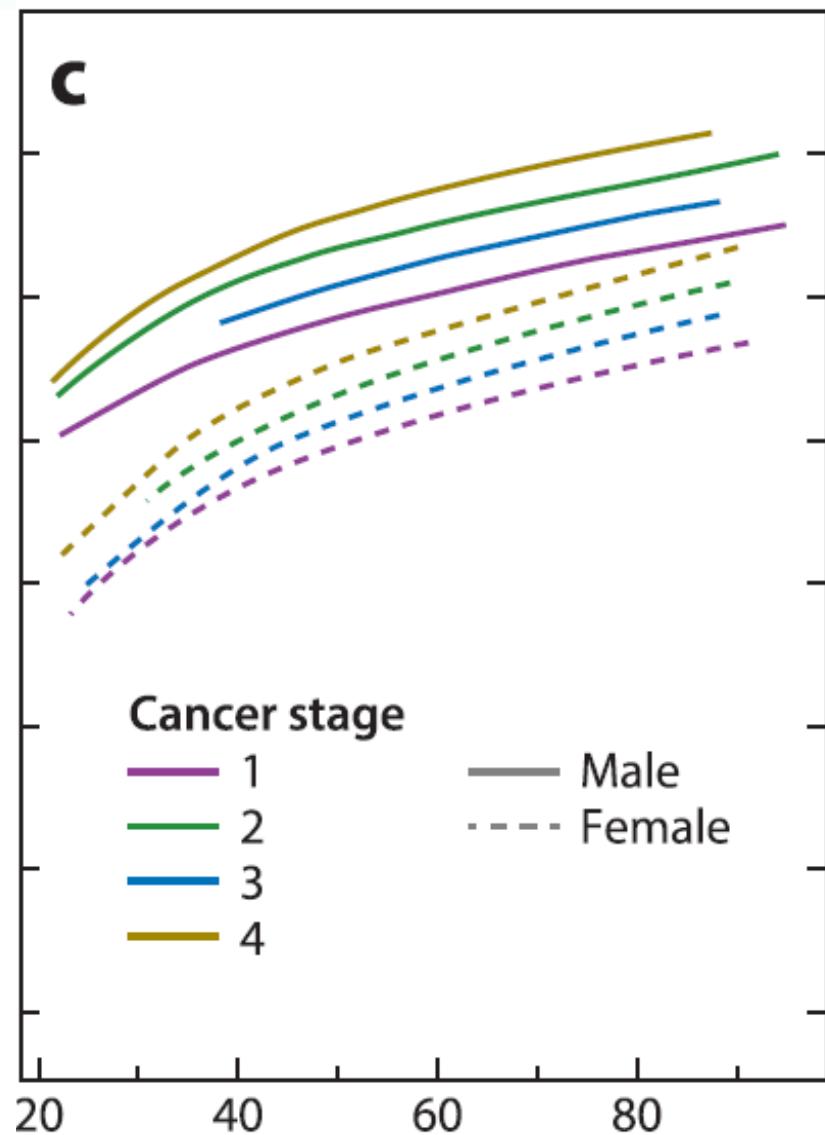
Cancer stage

- | | |
|---|--------|
| 1 | Male |
| 2 | Female |
| 3 | Male |
| 4 | Female |

Death



Progression -> Death



Older age is associated with higher hazards;
interestingly, while sex does not seem to play a role in disease progression .

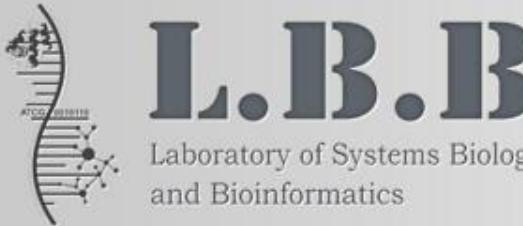


Simulation Settings					Maximum Likelihood Estimation			Deep Neural Network		
Setting	Sample Size	Frailty Variance (θ)	Log-Risk Function	Censoring Rate	h_1	h_2	h_3	h_1	h_2	h_3
1	1,000	0.50	Linear	0%	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.07 (0.05)	0.08 (0.08)	0.08 (0.05)
2	10,000	0.50	Linear	0%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.08 (0.07)	0.08 (0.05)	0.08 (0.07)
3	1,000	2.00	Linear	0%	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)	0.12 (0.07)	0.13 (0.07)	0.13 (0.09)
4	10,000	2.00	Linear	0%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.11 (0.06)	0.11 (0.08)	0.13 (0.10)
5	1,000	0.50	Non-Linear	0%	1.80 (0.33)	1.82 (0.39)	1.85 (0.34)	0.09 (0.05)	0.09 (0.04)	0.08 (0.04)
6	10,000	0.50	Non-Linear	0%	1.80 (0.13)	1.77 (0.13)	1.78 (0.11)	0.07 (0.03)	0.08 (0.03)	0.08 (0.05)
7	1,000	2.00	Non-Linear	0%	1.92 (0.53)	1.85 (0.54)	1.96 (0.53)	0.15 (0.05)	0.15 (0.06)	0.14 (0.05)
8	10,000	2.00	Non-Linear	0%	1.82 (0.17)	1.81 (0.18)	1.83 (0.18)	0.14 (0.04)	0.12 (0.03)	0.13 (0.06)
9	1,000	0.50	Linear	25%	0.01 (0.02)	0.02 (0.01)	0.02 (0.02)	0.10 (0.06)	0.10 (0.07)	0.13 (0.12)
10	10,000	0.50	Linear	25%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.12 (0.10)	0.12 (0.09)	0.12 (0.10)
11	1,000	2.00	Linear	25%	0.03 (0.02)	0.02 (0.02)	0.04 (0.03)	0.15 (0.10)	0.13 (0.09)	0.18 (0.12)
12	10,000	2.00	Linear	25%	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.14 (0.10)	0.12 (0.08)	0.14 (0.10)
13	1,000	0.50	Non-Linear	25%	1.96 (0.44)	2.01 (0.54)	2.24 (0.66)	0.10 (0.07)	0.10 (0.06)	0.10 (0.08)
14	10,000	0.50	Non-Linear	25%	1.95 (0.15)	1.91 (0.16)	2.16 (0.20)	0.07 (0.04)	0.09 (0.08)	0.09 (0.07)
15	1,000	2.00	Non-Linear	25%	2.06 (0.62)	1.92 (0.72)	2.25 (0.79)	0.15 (0.08)	0.15 (0.08)	0.13 (0.06)
16	10,000	2.00	Non-Linear	25%	1.88 (0.21)	1.88 (0.21)	2.04 (0.28)	0.10 (0.05)	0.11 (0.06)	0.11 (0.05)
17	1,000	0.50	Linear	50%	0.01 (0.02)	0.02 (0.02)	0.04 (0.03)	0.10 (0.07)	0.10 (0.06)	0.20 (0.15)
18	10,000	0.50	Linear	50%	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.10 (0.07)	0.11 (0.08)	0.17 (0.16)
19	1,000	2.00	Linear	50%	0.03 (0.03)	0.03 (0.02)	0.05 (0.05)	0.22 (0.13)	0.17 (0.13)	0.24 (0.17)
20	10,000	2.00	Linear	50%	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)	0.14 (0.09)	0.14 (0.10)	0.16 (0.14)
21	1,000	0.50	Non-Linear	50%	2.06 (0.50)	2.20 (0.72)	2.61 (1.00)	0.09 (0.06)	0.13 (0.13)	0.18 (0.14)
22	10,000	0.50	Non-Linear	50%	2.03 (0.21)	2.00 (0.18)	2.36 (0.25)	0.06 (0.03)	0.09 (0.08)	0.10 (0.09)
23	1,000	2.00	Non-Linear	50%	2.16 (0.76)	2.00 (0.72)	2.41 (0.91)	0.18 (0.10)	0.18 (0.09)	0.16 (0.10)
24	10,000	2.00	Non-Linear	50%	1.92 (0.25)	1.95 (0.23)	2.22 (0.38)	0.10 (0.05)	0.11 (0.06)	0.15 (0.13)



Method	Citation	Available Software
Classical Survival Analysis		
Cox Proportional Hazards Model Accelerated Failure Time Models Censored Quantile Regression Sub-Distribution Hazard Model Illness-Death Model	Cox (1972) Buckley & James (1979) Portnoy (2003) Fine & Gray (1999) Haneuse & Lee (2016)	https://cran.r-project.org/web/packages/survival/index.html https://cran.r-project.org/web/packages/survival/index.html https://cran.r-project.org/web/packages/quantreg/index.html https://cran.r-project.org/web/packages/cmprsk/index.html https://cran.r-project.org/web/packages/SemiCompRisks/index.html
Regularized Cox Models		
Ridge LASSO Elastic Net Adaptive LASSO SCAD Group LASSO Fused LASSO Graphical LASSO	Verweij & Van Houwelingen (1994) Tibshirani (1997) Wu (2012) Zhang & Lu (2007) Fan & Li (2002) Kim et al. (2012) Tibshirani et al. (2005) Friedman et al. (2008)	https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/gcvreg/index.html https://cran.r-project.org/web/packages/grpreg/index.html https://cran.r-project.org/web/packages/penalized/penalized.pdf -
Dantzig Selector		
AFT Models Cox Model	Li et al. (2014) Antoniadis et al. (2010)	http://www-personal.umich.edu/~yili/adsfxns.R -
Ultra High-Dimensional Feature Screening		
Sure Independence Screening Principled Sure Independent Screening Buckley-James Assisted Sure Screening Conditional Screening Forward Regression	Fan & Lv (2008) Fan & Song (2010) Zhao & Li (2012) Liu et al. (2020) Kang et al. (2017) Hong et al. (2019)	https://cran.r-project.org/web/packages/SIS/index.html https://cran.r-project.org/web/packages/SIS/index.html http://faculty.washington.edu/acock/software.html - https://github.com/younghhk/software/blob/master/CS.R -
Inferential Methods		
Selection-Assisted Partial Regression and Smoothing Fused High-Dimensional Censored Quantile Regression	Fei & Li (2021) Fei et al. (2021)	https://github.com/feizhe/SPARES https://github.com/feizhe/HDCQR_Paper
Support Vector Machines		
Rank-Based Approach Regression Approach Hybrid Approach	Van Belle et al. (2007) Shivaswamy et al. (2007) Van Belle et al. (2011) Pöhlsterl et al. (2015)	https://cran.r-project.org/web/packages/survivalsvm/index.html https://cran.r-project.org/web/packages/survivalsvm/index.html https://cran.r-project.org/web/packages/survivalsvm/index.html https://cran.r-project.org/web/packages/survivalsvm/index.html
Tree-Based Methods		
Log-Rank Based Likelihood-Based	Ciampi et al. (1986) Ciampi et al. (1987)	https://cran.r-project.org/web/packages/rpart/index.html https://cran.r-project.org/web/packages/rpart/index.html
Ensemble Learners		
Bootstrap Aggregation Gradient Boosting Random Survival Forests Censoring Unbiased Regression Trees	Hothorn et al. (2004) Hothorn et al. (2006) Ishwaran et al. (2008) Ishwaran et al. (2011) Ishwaran & Lu (2019) Steingrimsson et al. (2019)	https://cran.r-project.org/web/packages/ipred/index.html https://cran.r-project.org/web/packages/gbm/index.html https://cran.r-project.org/web/packages/randomForestSRC/index.html https://cran.r-project.org/web/packages/randomForestSRC/index.html https://cran.r-project.org/web/packages/randomForestSRC/index.html https://cran.r-project.org/web/packages/randomForest/index.html
Deep Learning		
DeepSurv DNNSurv	Katzman et al. (2018) Zhao & Feng (2020)	https://cran.r-project.org/web/packages/survivalmodels/index.html https://cran.r-project.org/web/packages/survivalmodels/index.html
Competing Risks		
DeepHit Dynamic DeepHit DeepCompete Hierarchical Multi-Event	Lee et al. (2018) Lee et al. (2019) Aastha & Liu (2020) Tjandra et al. (2021)	https://cran.r-project.org/web/packages/survivalmodels/index.html https://github.com/chl8856/Dynamic-DeepHit - -
Semi-Competing Risks		
Penalized Estimation Deep Learning	Reeder et al. (2022) <i>Proposed</i>	- -

Method	Available Software
Classical Survival Analysis <ul style="list-style-type: none"> Cox Proportional Hazards Model Accelerated Failure Time Models Censored Quantile Regression Sub-Distribution Hazard Model Illness-Death Model 	https://cran.r-project.org/web/packages/survival/index.html https://cran.r-project.org/web/packages/survival/index.html https://cran.r-project.org/web/packages/quantreg/index.html https://cran.r-project.org/web/packages/cmprsk/index.html https://cran.r-project.org/web/packages/SemiCompRisks/index.html
Regularized Cox Models <ul style="list-style-type: none"> Ridge LASSO Elastic Net Adaptive LASSO SCAD Group LASSO Fused LASSO Graphical LASSO 	https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/glmnet/glmnet.pdf https://cran.r-project.org/web/packages/ncvreg/index.html https://cran.r-project.org/web/packages/grpreg/index.html https://cran.r-project.org/web/packages/penalized/penalized.pdf -
Dantzig Selector <ul style="list-style-type: none"> AFT Models Cox Model 	http://www-personal.umich.edu/~yili/adsfxns.R -



Laboratory of Systems Biology
and Bioinformatics



Siamak-salimy

Edit profile

16 followers · 25 following

Achievements



Beta Send feedback

Pinned

Customize your pins

[SiamakSalimy](#) Public ⋮

☆ 1

[SingleCellWorkshop](#) Public ⋮

● R

[High-Dimensional-Survival-Analysis](#) Public ⋮

[ConditionalCalculator](#) Public ⋮

Forked from [LBBSoft/ConditionalCalculator](#)
A Deep Learning-Based Framework for Predicting Survival-Associated Groups in Colon Cancer by Integrating Multi-Omics and Clinical Data.
● ASP.NET

[Bulk-RNA-seq](#) Public ⋮

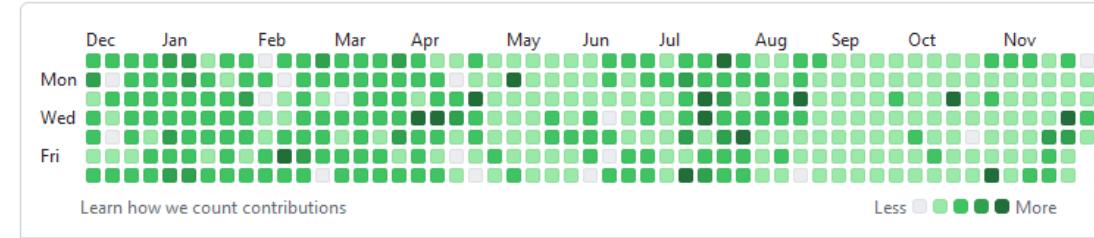
[Biological_DataBase](#) Public ⋮

● R

1,072 contributions in the last year

Contribution settings ▾

2023



<https://github.com/Siamak-salimy/High-Dimensional-Survival-Analysis>



L.B.B

Laboratory of Systems Biology
and Bioinformatics

Thanks



Criteria	SCAD	LASSO	MCP
Penalty Function	Smoothly encourages sparsity with a kink	Encourages sparsity abruptly	Gradually encourages sparsity
Variable Selection	Soft selection of variables; fewer false positives	Prone to false positives	Fewer false positives, intermediate between SCAD and LASSO
Coefficient Shrinkage	Gradual, avoids abrupt shrinkage	Abruptly shrinks coefficients	Gradual shrinkage, intermediate behavior
Zero Coefficients	Tends to produce exact zero coefficients	Tends to produce exact zero coefficients	Fewer exact zero coefficients, intermediate behavior
Threshold Behavior	Kink in penalty function controls shrinkage	Lambda parameter controls shrinkage	Lambda parameter controls shrinkage
Performance	May outperform LASSO in specific scenarios	Widely used but has limitations	Balances between LASSO and SCAD
Mathematical Formulation	Complex penalty function	Simple penalty function	Intermediate complexity in penalty function
Tuning Parameters	Parameters controlling kink and shrinkage	Lambda parameter controlling shrinkage	Tuning parameter controlling sparsity