Randomized Stepwise Regression

A Modification of Stepwise Methods in Generalized Linear Models and its Application on Sepsis Data

Marcus Vollmer

Institute of Bioinformatics, University Medicine Greifswald

Reisensburg b. Günzburg
10 July 2018

Universitätsmedizin GREIFSWALD

Outline

1. Purpose & State of the Art
2. Stepwise Regression
3. Randomized Stepwise Regression
4. Performance on real data

Purpose & State of the Art

Purpose

Fitting a statistical model for time to event data can be challenging in many instances. Our analysis is influenced through **censored data**, **missing values** and **correlation**. In many cases we have many variables P measured in few cases N. Therefore, we need to handle the problem of **overfitting**.

Purpose

Fitting a statistical model for time to event data can be challenging in many instances.
Our analysis is influenced through **censored data**, **missing values** and **correlation**. In many cases we have many variables P measured in few cases N. Therefore, we need to handle the problem of **overfitting**.

Events Per Variable (EPV) for regression analysis

- Harrell et al. suggested a minimum of 10 to 20 EPV

  Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–387.

- Peduzzi et al. performed a simulation study, and suggested that at least 10 EPV are needed to maintain the validity of the model

  Peduzzi et al., Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates, Journal of Clinical Epidemiology, Volume 48, Issue 12, December 1995, Pages 1503-1510

Cox Regression

```
model = coxph(Surv(survival_days90, survival_event90) ~ ., data=ds)
stargazer(model)
```

		β coefficient	$\sigma(\beta)$
Sex	male	−0.198	(0.309)
ICU admission reason	surgical emergency	0.180	(0.549)
	surgical planed	0.781	(0.642)
Lactate level	2-4 mmol/L	0.622	(0.451)
	>4 mmol/L	1.079**	(0.530)
⋮			
Betablocker treatment	continued	−0.430	(0.334)
Observations	176		
Log Likelihood	−343.658		
Score (Logrank) Test	113.940*** (df = 49)		
Note:	*p<0.1; **p<0.05; ***p<0.01		

Cox Regression

```r
model = coxph(Surv(survival_days90, survival_event90) ~ ., data=ds)
stargazer(model)
```

		β coefficient	$\sigma(\beta)$
Sex	male	−0.198	(0.309)
ICU admission reason	surgical emergency	0.180	(0.549)
	surgical planed	0.781	(0.642)
Lactate level	2-4 mmol/L	0.622	(0.451)
	>4 mmol/L	1.079**	(0.530)
⋮			
Betablocker treatment	continued	−0.430	(0.334)
Observations	176		
Log Likelihood	−343.658		
Score (Logrank) Test	113.940*** (df = 49)		
Note:	*p<0.1; **p<0.05; ***p<0.01		

Overfitted!

Shrinkage Procedures

Regularized parameter estimation

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \left(-\frac{2}{N}\ell(\beta) + \alpha\lambda \sum_{p=1}^{P} |\beta_p| + \frac{1-\alpha}{2}\lambda \sum_{p=1}^{P} \beta_p^2 \right)$$

- LASSO: $\alpha = 1$
- Ridge: $\alpha = 0$
- Elastic Net: $0 < \alpha < 1$

Shrinkage Procedures

Regularized parameter estimation

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \left(-\frac{2}{N}\ell(\beta) + \alpha\lambda \sum_{p=1}^{P} |\beta_p|\omega_p + \frac{1-\alpha}{2}\lambda \sum_{p=1}^{P} \beta_p^2 \omega_p \right)$$

- LASSO: $\alpha = 1$
- Ridge: $\alpha = 0$
- Elastic Net: $0 < \alpha < 1$
- Differential shrinkage: $\omega_p \neq 1$

Shrinkage Procedures

Regularized parameter estimation

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \left(-\frac{2}{N}\ell(\beta) + \alpha\lambda \sum_{p=1}^{P} |\beta_p| \omega_p + \frac{1-\alpha}{2}\lambda \sum_{p=1}^{P} \beta_p^2 \omega_p \right)$$

- LASSO: $\alpha = 1$
- Ridge: $\alpha = 0$
- Elastic Net: $0 < \alpha < 1$
- Differential shrinkage: $\omega_p \neq 1$

- Adaptive LASSO [1]: $\omega_p = \left|\hat{\beta}_p\right|^{-\gamma}, \gamma > 0$
- Adaptive Elastic Net [2]: $\omega_p = \left|\hat{\beta}_p\right|^{-\gamma}, \gamma > 0$

LASSO in R

```
library("glmnet")

x = model.matrix( ~ ., ds[,-(c(1:2))])
y = Surv(ds$survival_days90, ds$survival_event90)

# 10-fold cross-validation fit
  cv.fit.lasso = cv.glmnet(x, y, family="cox", alpha=1)

# coefficients at minimum mean cross-validated error
  c = coef(cv.fit.lasso, s="lambda.min")
  colnames(x[,rowSums(c!=0)>0])
```

```
[1] "ChronischeErkrankungen2" "LeukozytenDiskret3"
[3] "SAPSIIScore"             "APACHEIIScore"
[5] "Erste24hLaktatDiskret3"  "EKerste24hDiskret2"
[7] "BBGruppe2"
```

Criteria of model selection

Balancing goodness of fit with simplicity

Some useful criteria:

1. Adjusted R^2
2. Mallows's C_p
3. Cross-validation
4. Akaike information criterion (AIC, AICc, CAIC)
5. Bayesian information criterion (BIC, BICc)

Model penalty

$\mathrm{BIC} = -2 \log L(\hat{\theta}) + k \log N,$ k estimated parameters, sample size N

Criteria of model selection

Balancing goodness of fit with simplicity

Some useful criteria:

1. Adjusted R^2
2. Mallows's C_p
3. Cross-validation
4. Akaike information criterion (AIC, AICc, CAIC)
5. Bayesian information criterion (BIC, BICc)

Model penalty

$$\hat{\beta} = \arg\min_{\beta} \left(-2\ell(\beta) + k \log N\right)$$

Criteria of model selection

Balancing goodness of fit with simplicity

Some useful criteria:

1. Adjusted R^2
2. Mallows's C_p
3. Cross-validation
4. Akaike information criterion (AIC, AICc, CAIC)
5. Bayesian information criterion (BIC, BICc)

Model penalty

$$\hat{\beta} = \arg\min_{\beta} \left(-2\ell(\beta) + \log N \sum_{p=1}^{P} \mathbb{1}_{\beta_p \neq 0} \right)$$

Universitätsmedizin GREIFSWALD

All subset regression

Computing all possible regression formulas and estimate (robustly) the coefficients of each model. The final model will be selected according to model criteria.

- Assume 30 possible predictor variables. There exists $2^{30} = 1,073,741,824$ subsets without interaction terms.
- Doable?

All subset regression

Computing all possible regression formulas and estimate (robustly) the coefficients of each model. The final model will be selected according to model criteria.

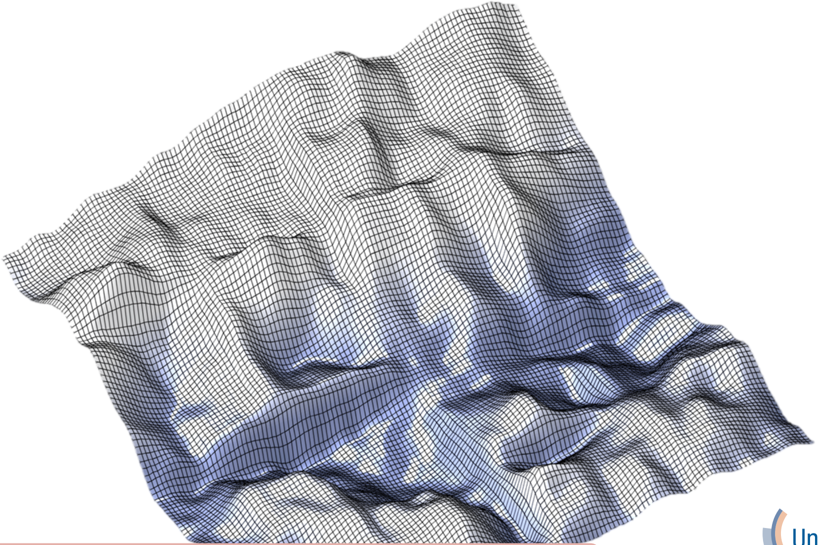
- Assume 30 possible predictor variables. There exists $2^{30} = 1,073,741,824$ subsets without interaction terms.
- Doable? Avg. 100 ms computing time for each model

1242.8 CPU days **CPU power > Energy > Costs!**

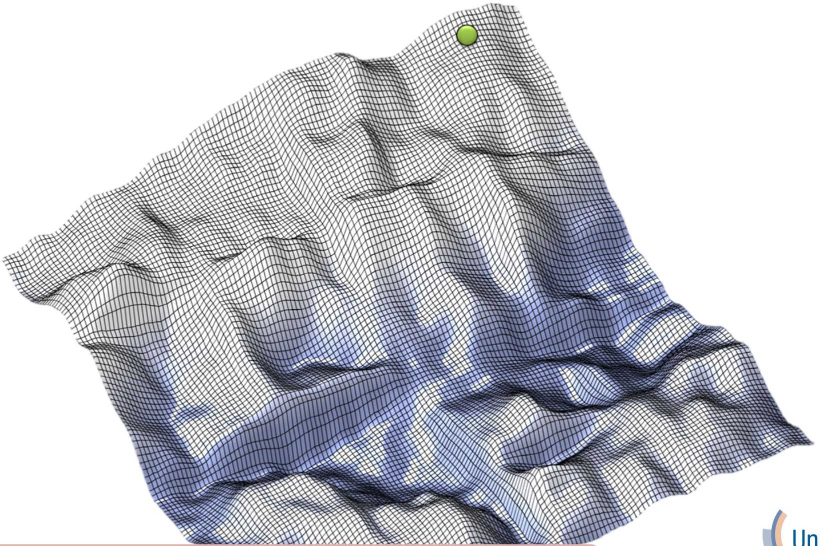
Universitätsmedizin GREIFSWALD

Stepwise Regression

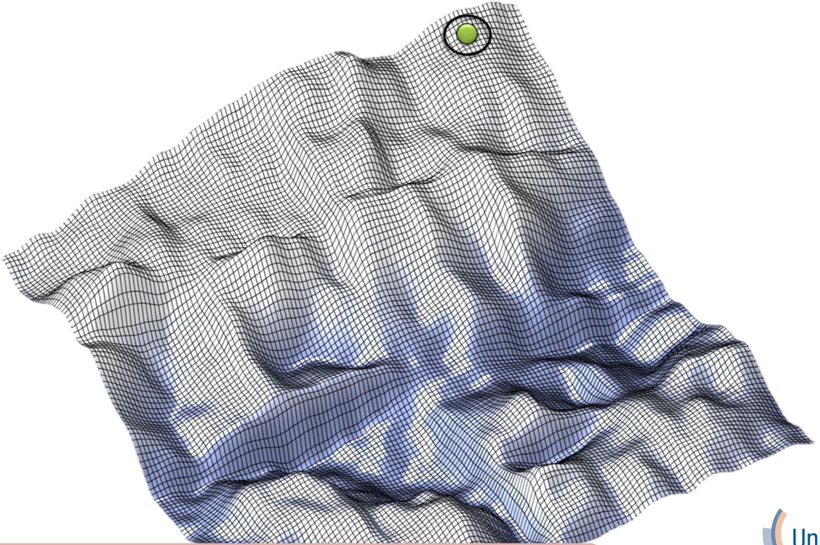
Classical Stepwise Regression



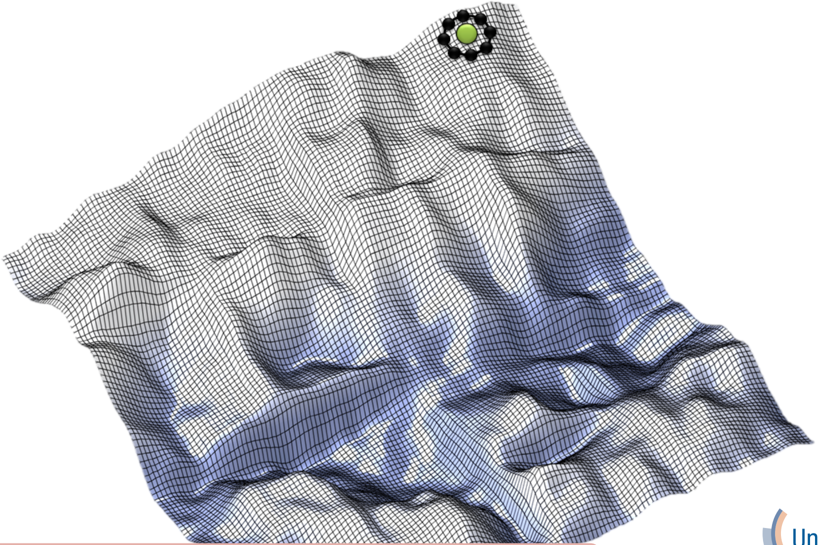
Classical Stepwise Regression



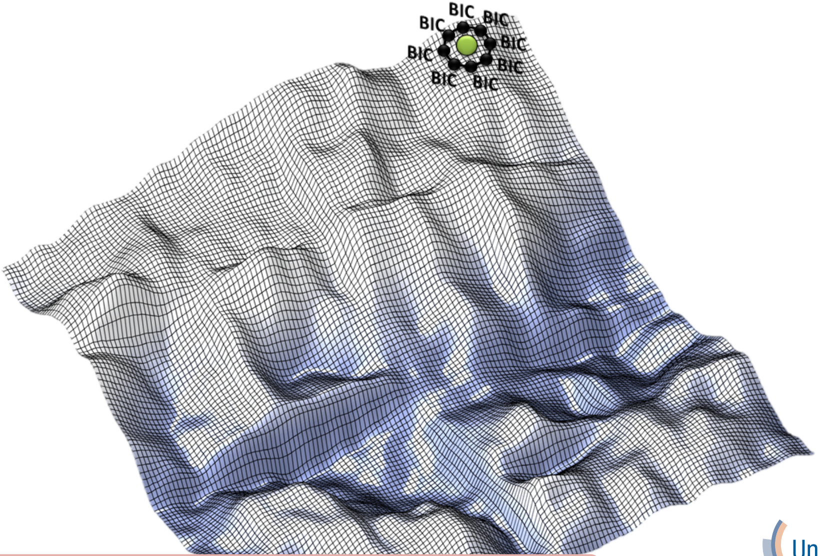
Classical Stepwise Regression



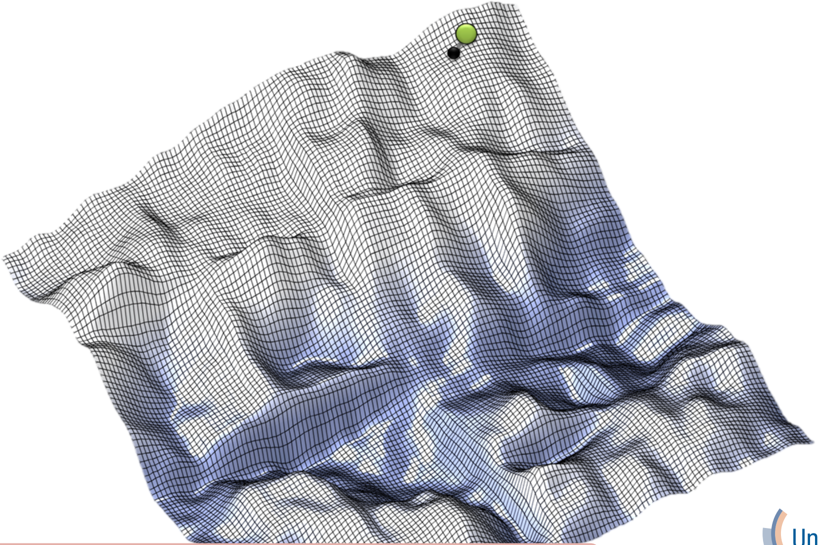
Classical Stepwise Regression



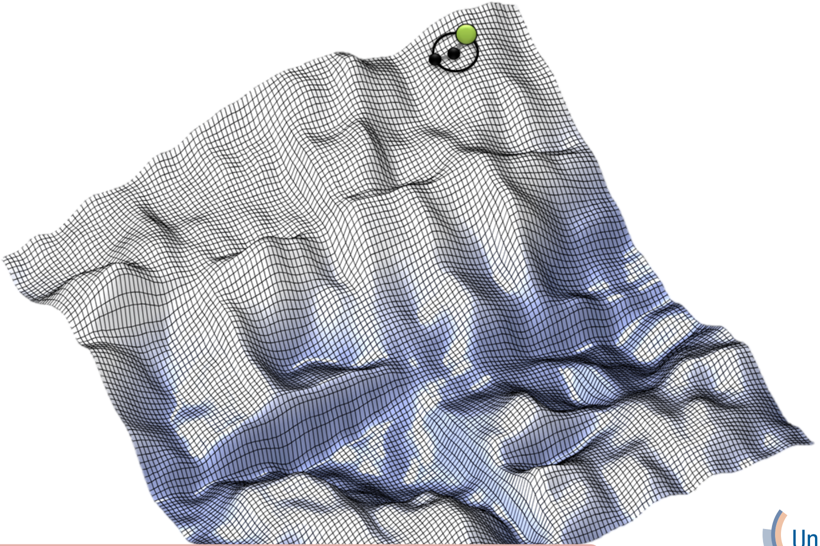
Classical Stepwise Regression



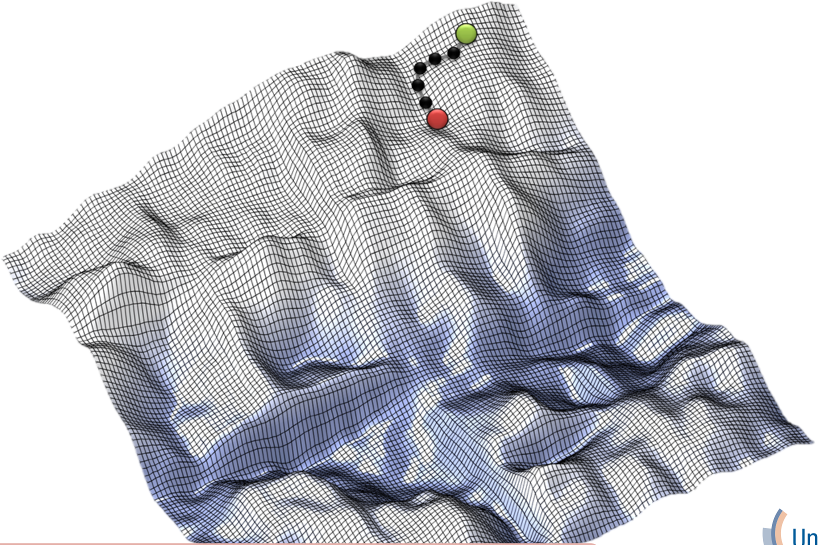
Classical Stepwise Regression



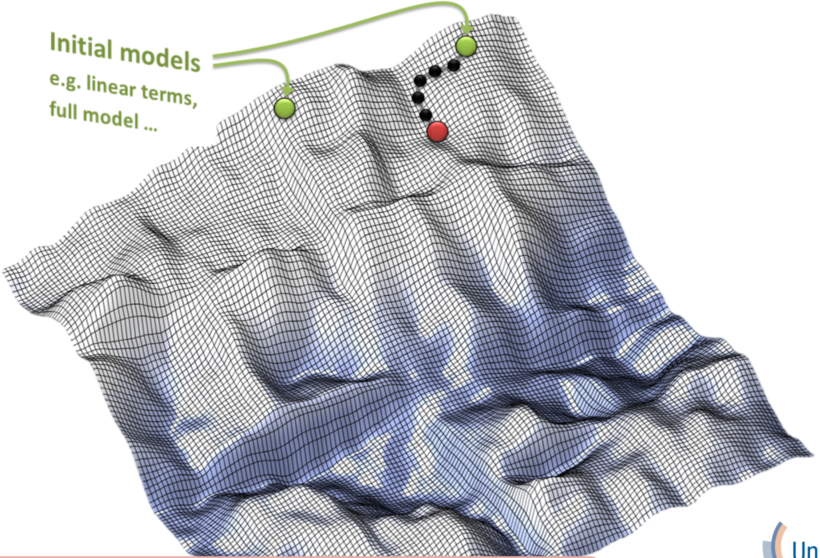
Classical Stepwise Regression



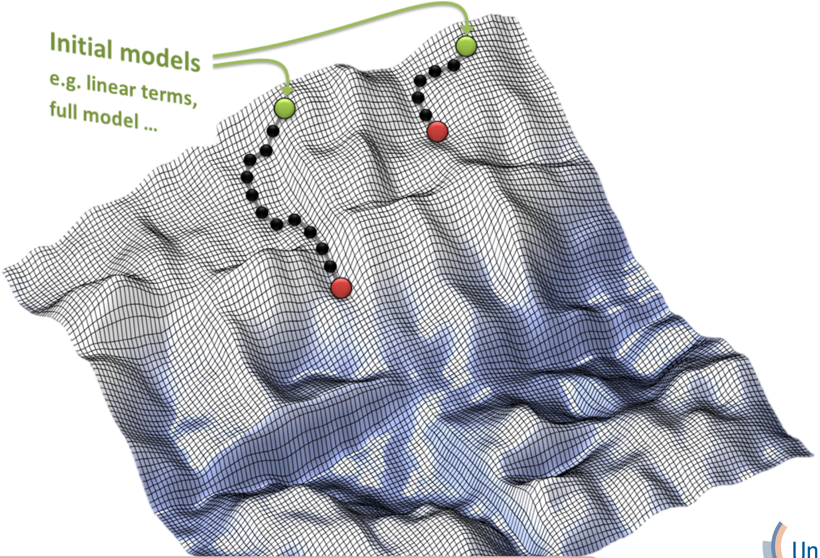
Classical Stepwise Regression



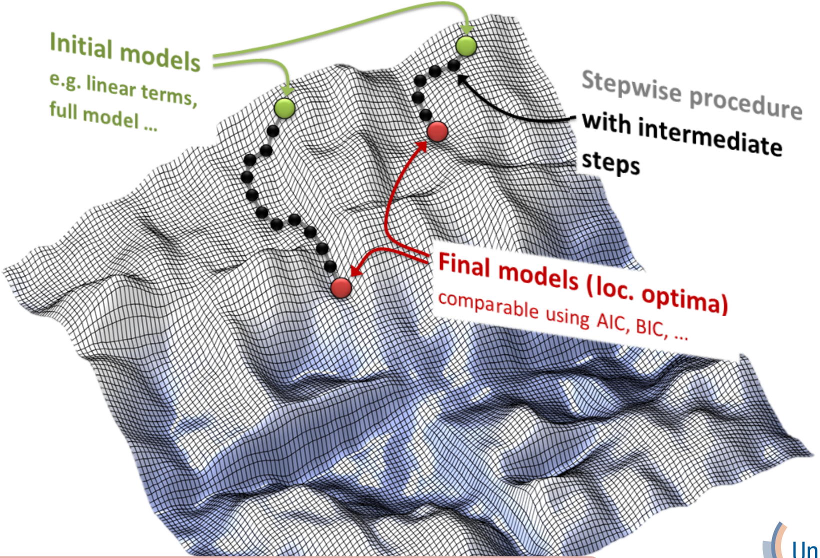
Classical Stepwise Regression



Classical Stepwise Regression

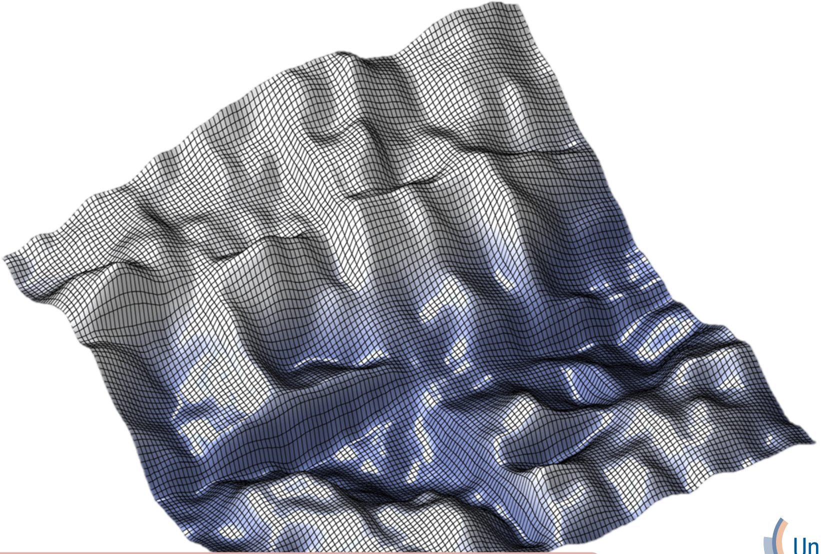


Classical Stepwise Regression



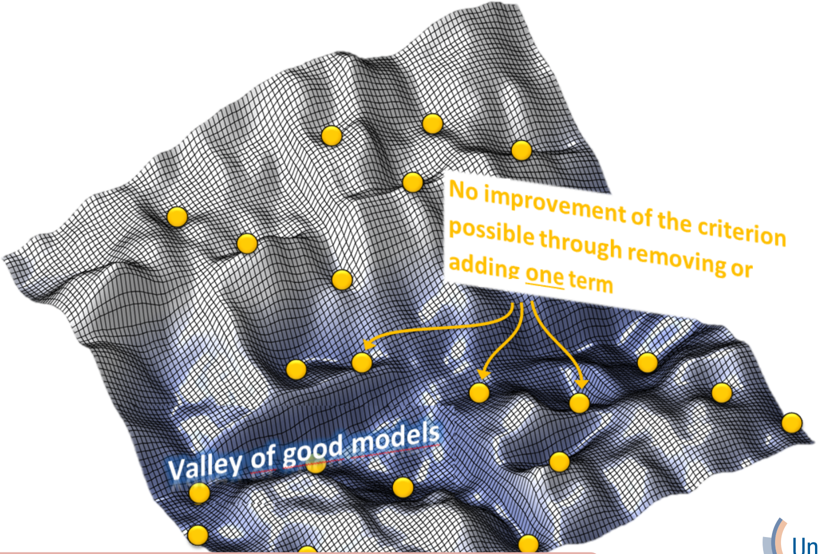
Universitätsmedizin GREIFSWALD

Model mountains with local optima



Universitätsmedizin GREIFSWALD

Model mountains with local optima

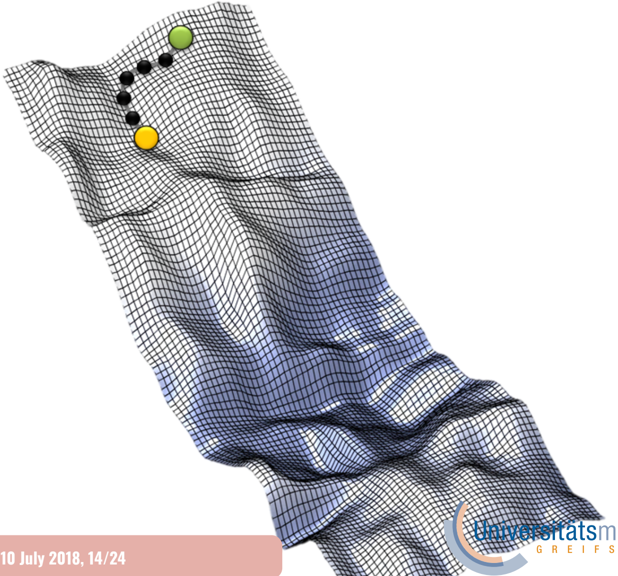


Universitätsmedizin
GREIFSWALD

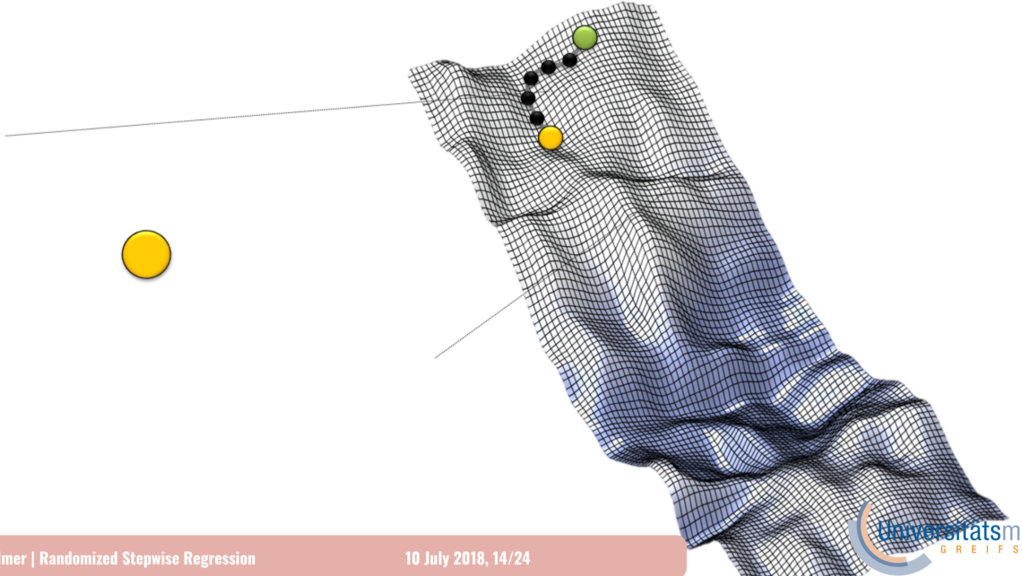
Randomized Stepwise Regression

Universitätsmedizin GREIFSWALD

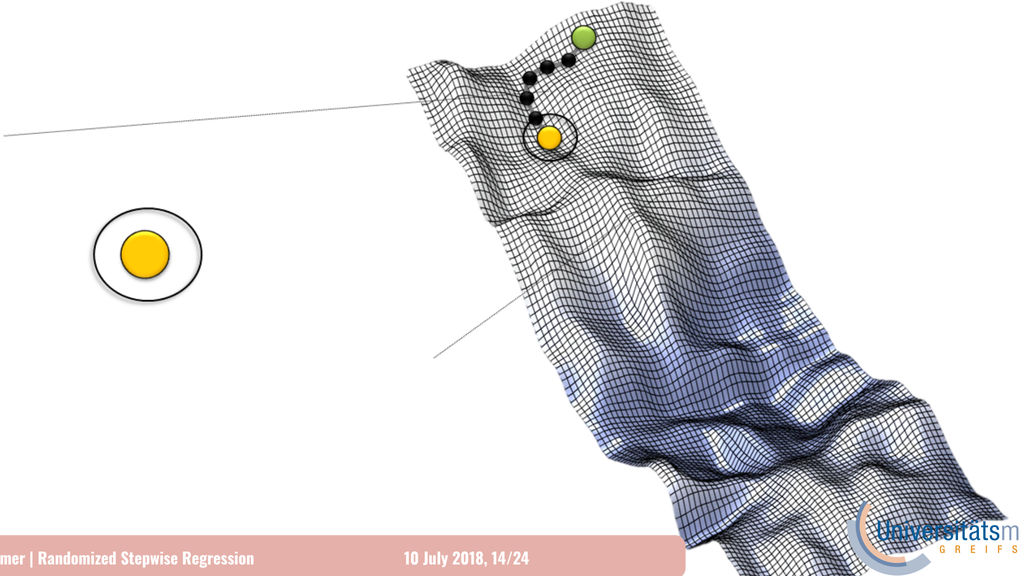
Randomized model selection



Randomized model selection

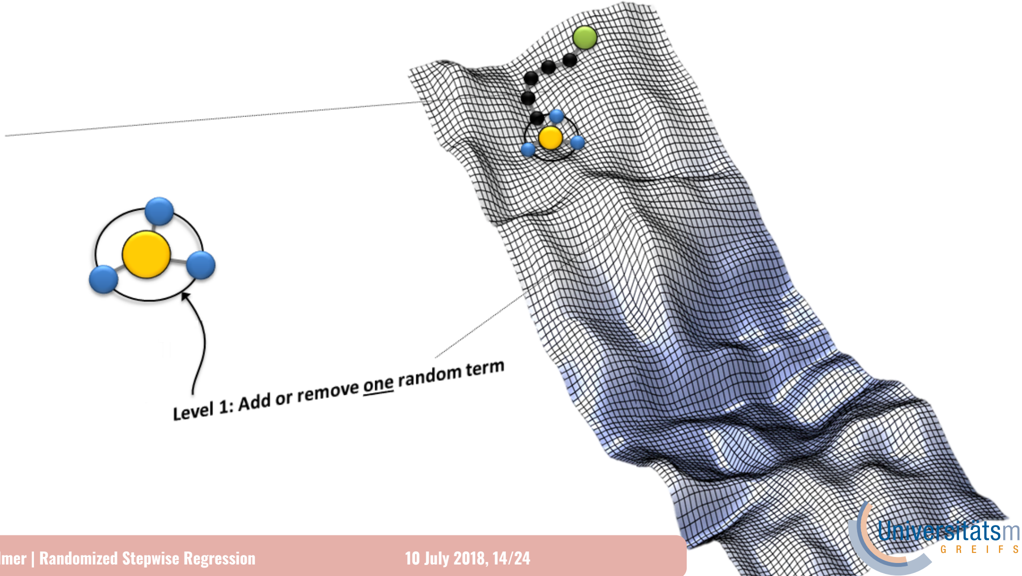


Randomized model selection

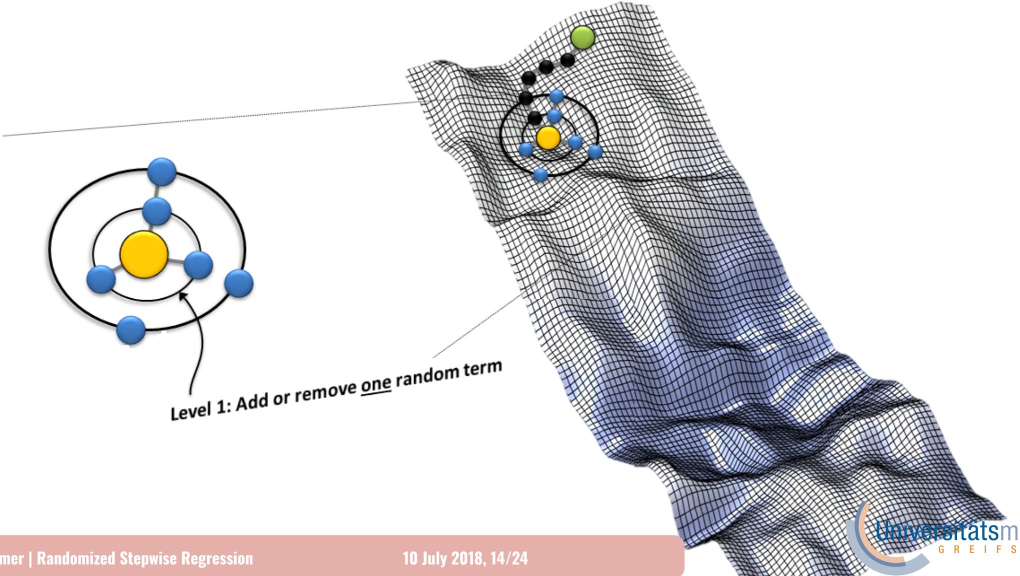


Universitätsmedizin GREIFSWALD

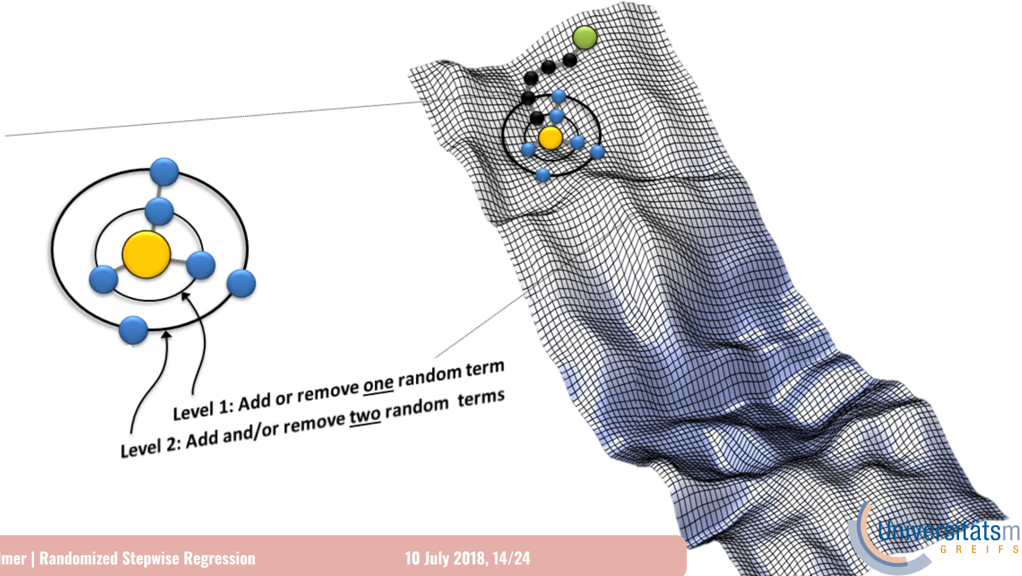
Randomized model selection



Randomized model selection

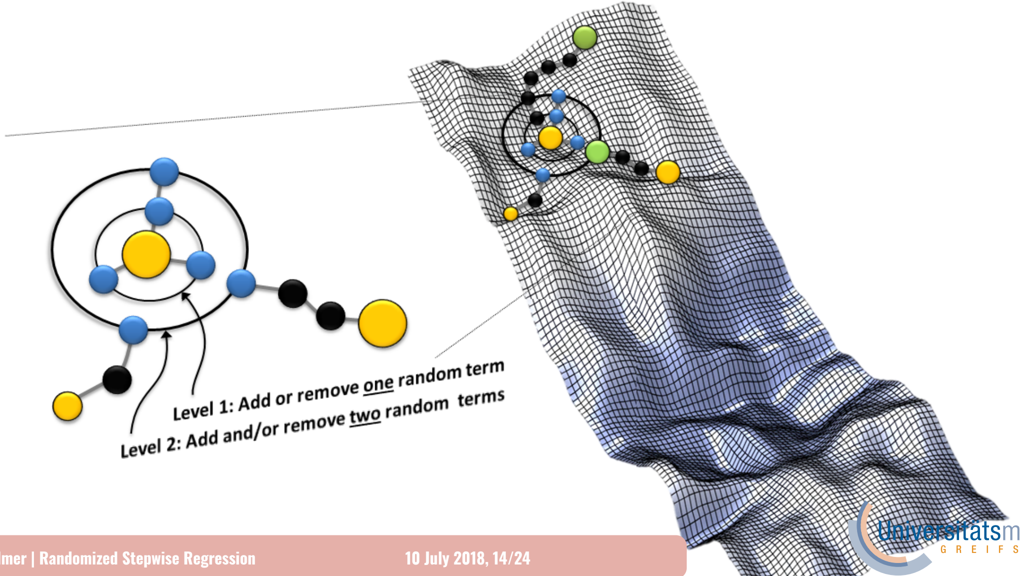


Randomized model selection

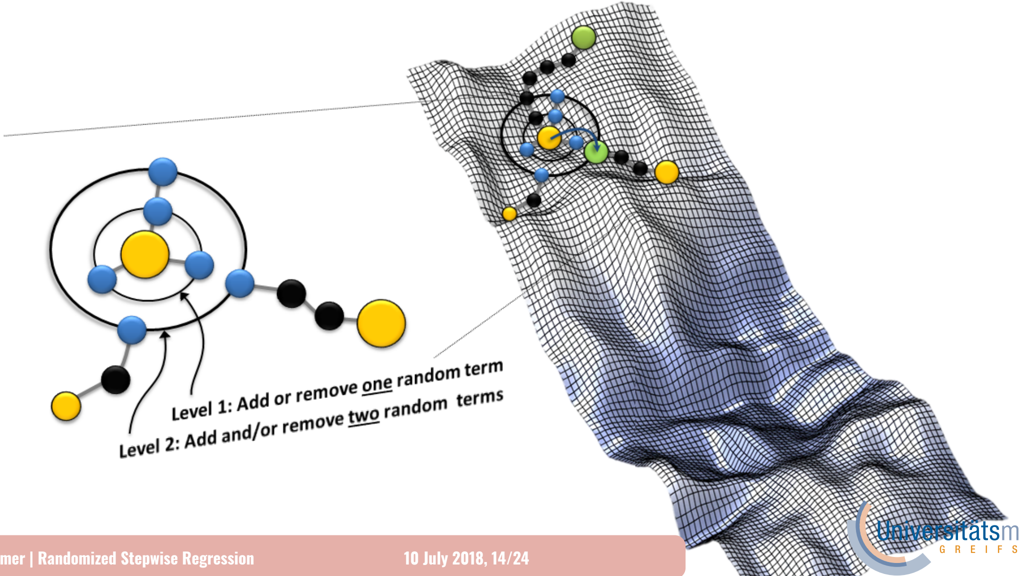


Universitätsmedizin GREIFSWALD

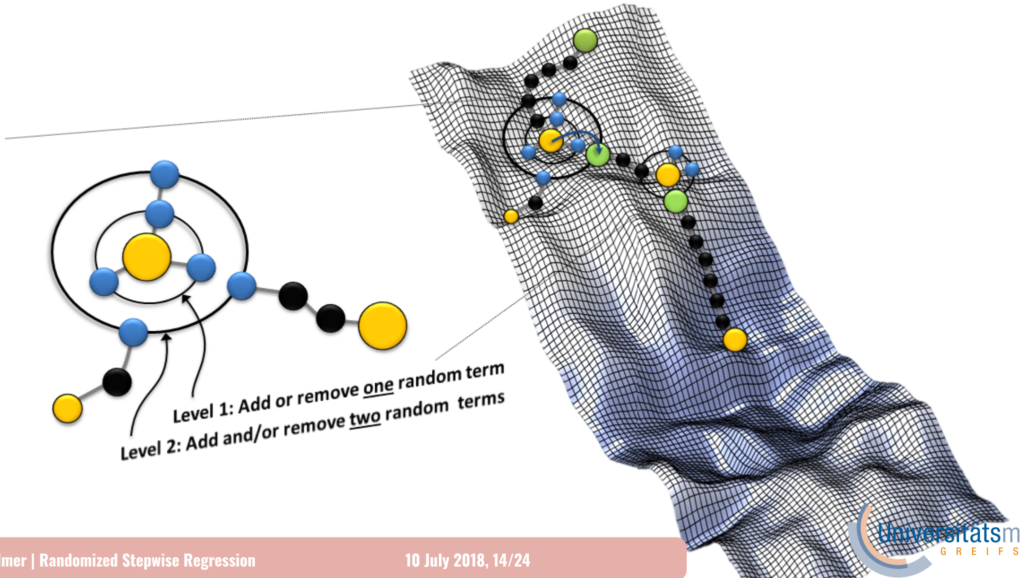
Randomized model selection



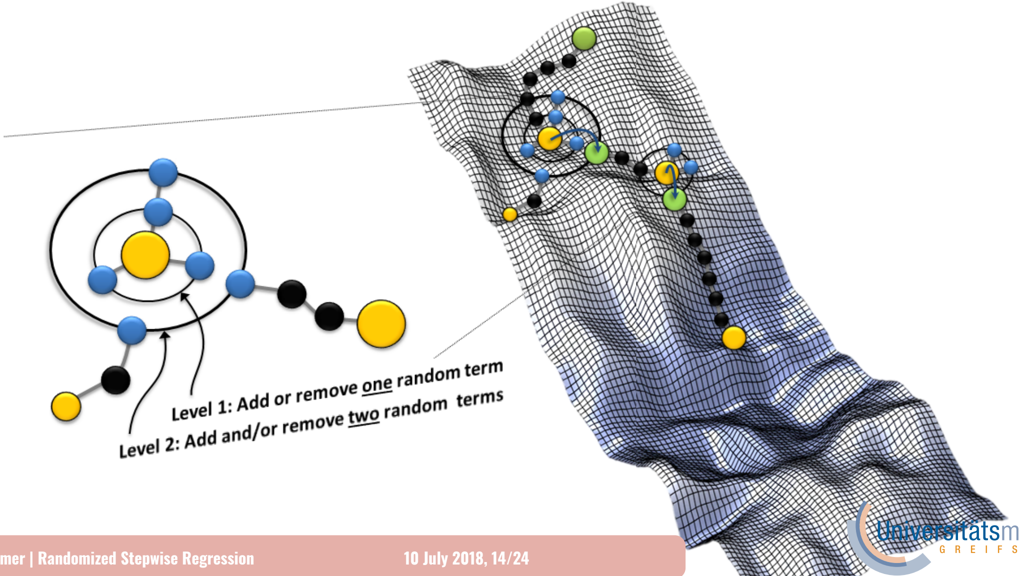
Randomized model selection



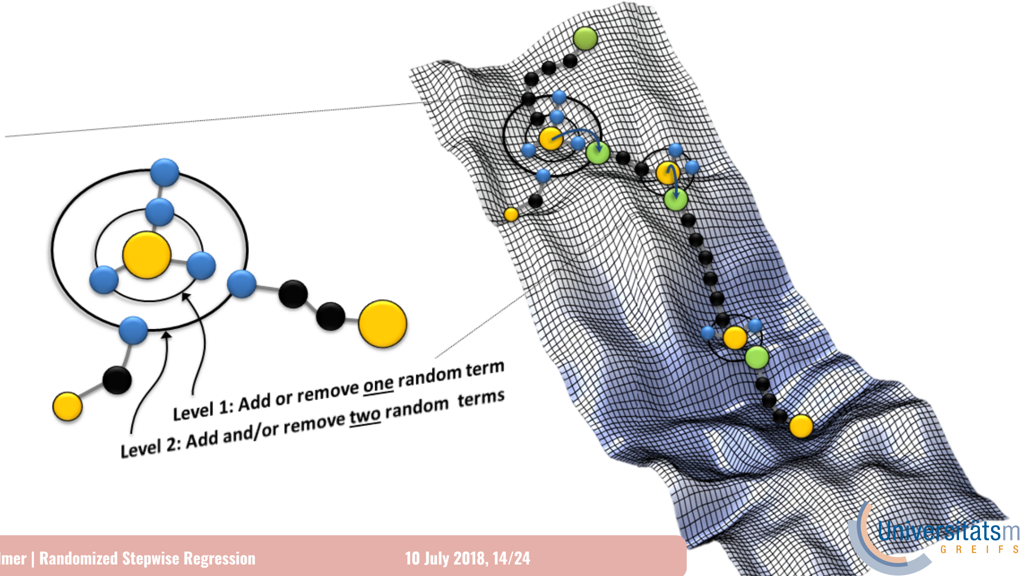
Randomized model selection



Randomized model selection

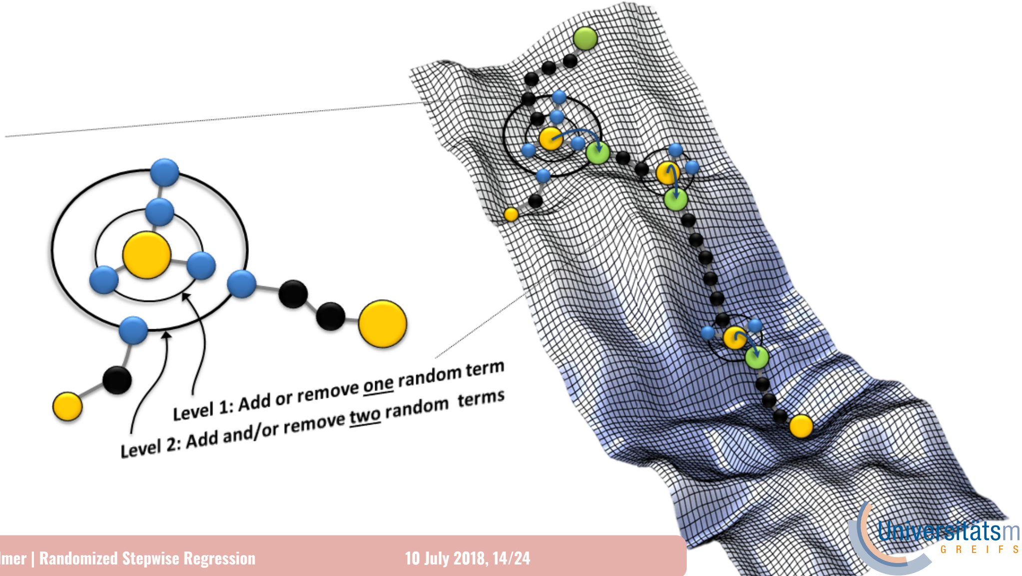


Randomized model selection



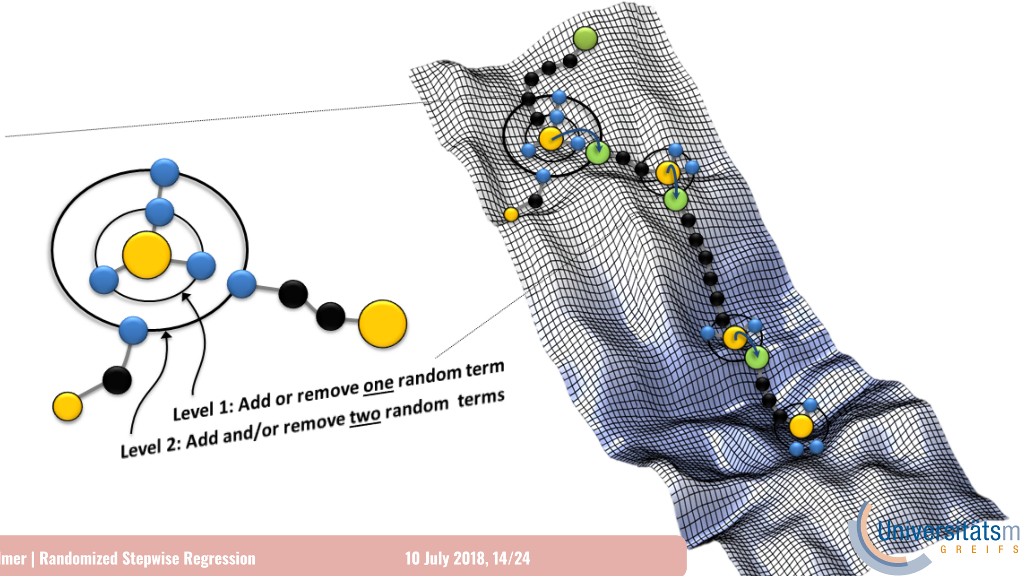
Universitätsmedizin GREIFSWALD

Randomized model selection

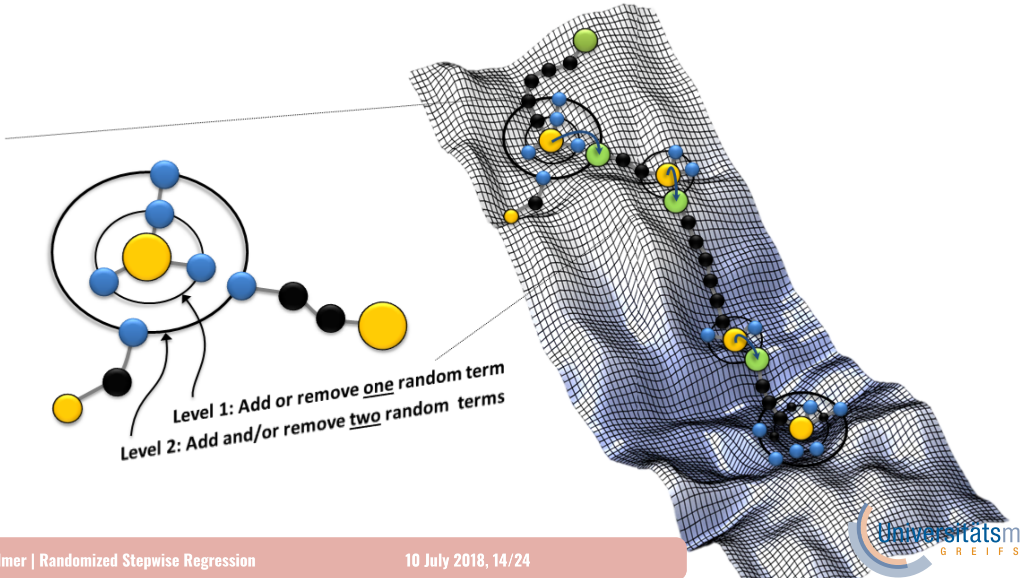


Universitätsmedizin GREIFSWALD

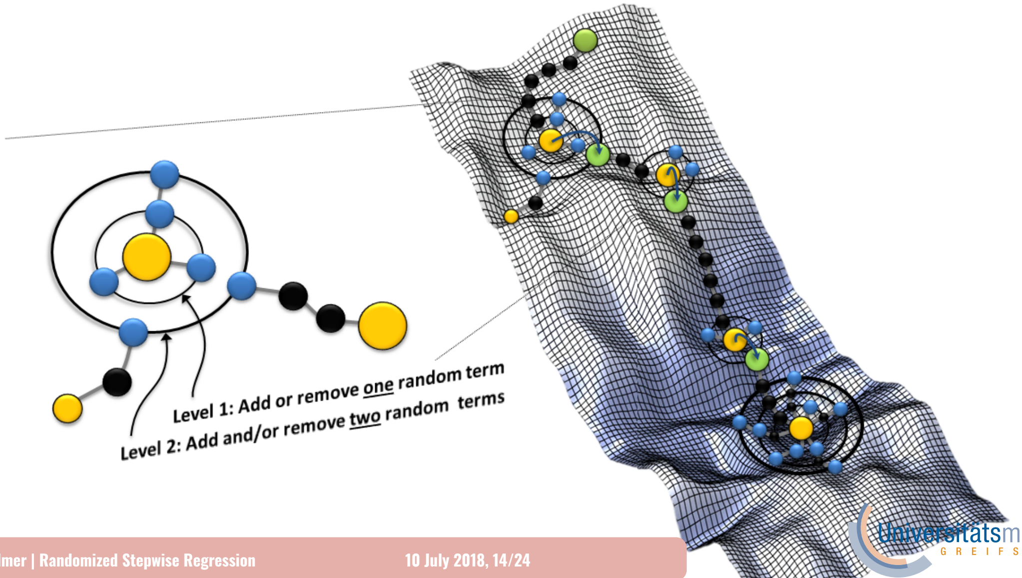
Randomized model selection



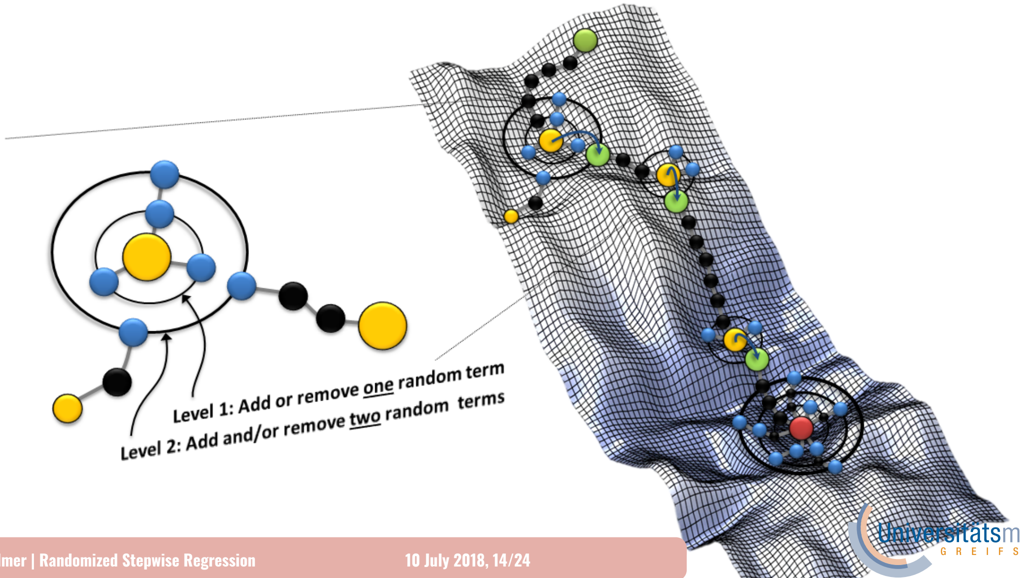
Randomized model selection



Randomized model selection



Randomized model selection



Randomized Stepwise Regression

Parameter set

- Initial model
  - fixed
  - arbitrary (proper distance)
- Specification
  - of model selection criteria
  - of the number of random models at each level
  - of a ratio of add and remove terms
  - of the maximal step length at each level
  - of an abort criterion (max. level)

> Hyperparameter optimization

> Adaptive parameter setting

Performance on real data

Data origin

The SepsisDialog

- Project at University Medicine Greifswald since 2008 [3]
- Improvement of diagnosis and therapy by
  1. awareness of sepsis
  2. advanced sepsis prophylaxis (e.g. hygiene)
  3. training of sepsis definition
  4. improvement of diagnostics
  5. improvement of primary treatment
     - early antibiotic administration
     - immediate rehabilitation of the source of infection
     - taking smears and blood cultures
     - fast stabilization of the circular system
- Professional training of hospital medical staff an nurses

Universitätsmedizin GREIFSWALD

Data origin

- $N = 793$ patients with septic shock or severe sepsis (SEP-1) from surgical intensive care unit (<4 missing values)
- Y: 90-day survival
- $P = 89$ predictor variables:

Onset condition

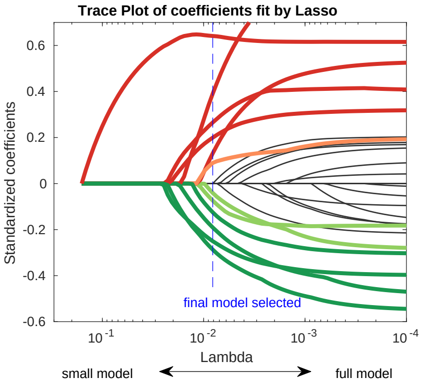
- 'Sex'
- 'Age'
- 'Sepsis severeness'
- 'APACHEII Score','SAPSII Score'
- 'Lactate level', serum blood parameters
- 'Preexisting antibiotic administration'
- chronic diseases ...

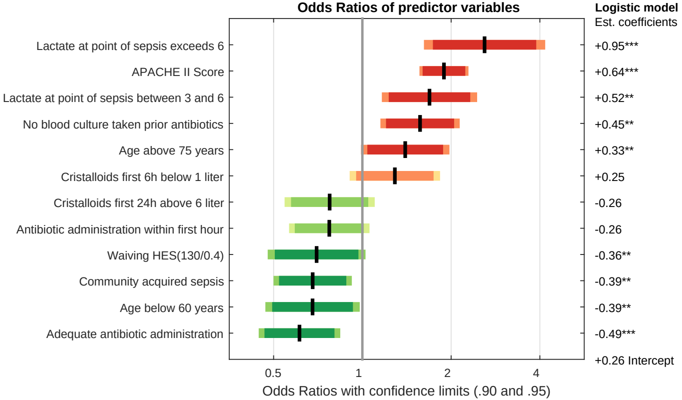
Treatment

- 'Appropriate antibiotic administration'
- 'Time to adequate antibiotic adm.'
- 'Smear test from source of infection'
- 'Crystalloid infusion first 6h'
- 'Antimycotics'
- other medication ...

Universitätsmedizin GREIFSWALD

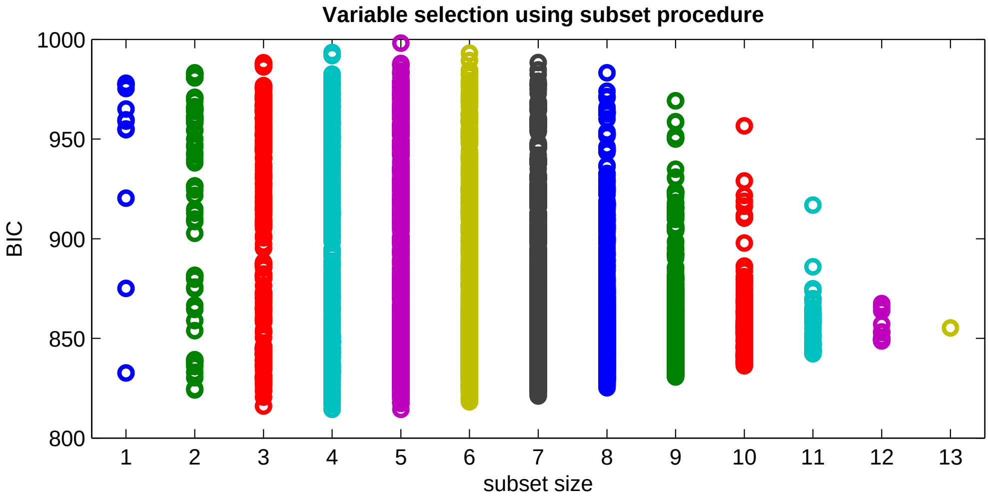
LASSO resuts



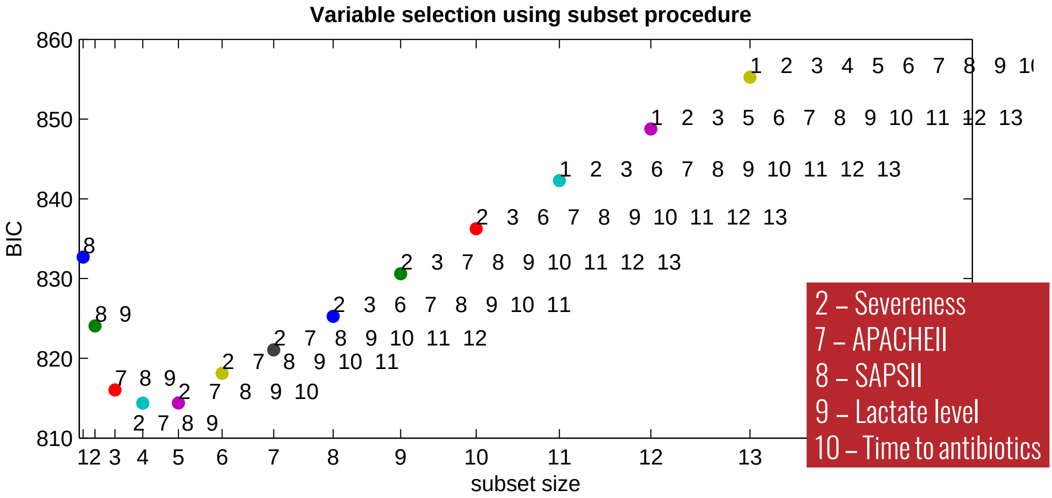


Universitätsmedizin GREIFSWALD

All subset procedure



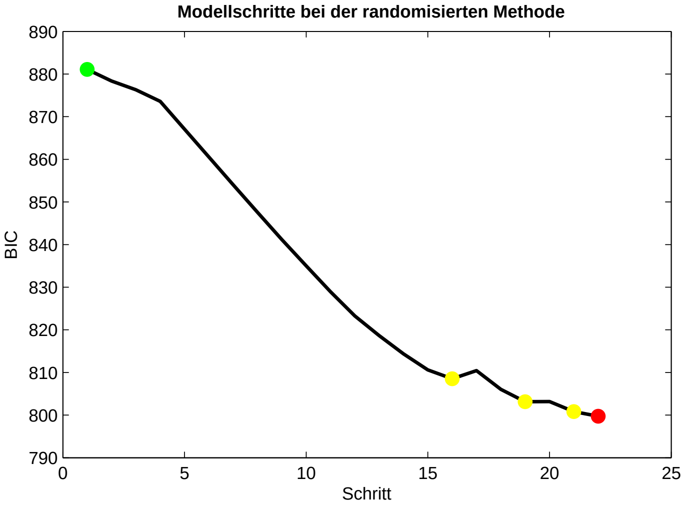
All subset procedure



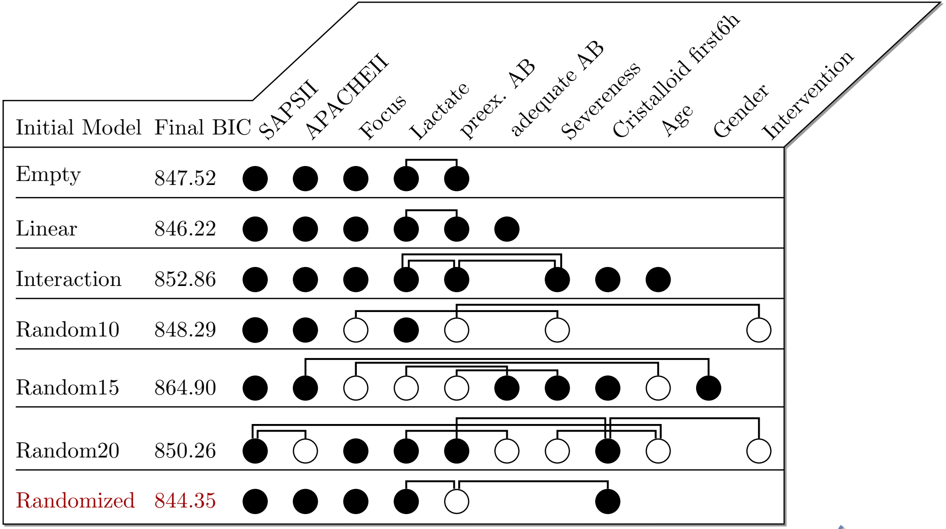
2 – Severeness
7 – APACHEII
8 – SAPSII
9 – Lactate level
10 – Time to antibiotics

Universitätsmedizin GREIFSWALD

Model steps of randomized model selection



Model comparison



Universitätsmedizin GREIFSWALD

Summary

Randomized Stepwise Regression

- The randomized strategy of the stepwise regression is a new model building strategy in generalized linear models
- Procedure guarantees a better or equivalent model compared to the classical approach, when starting with the same initial model
- Run-time is moderate (depending on parameter set)
- Finding of the optimal model not guaranteed (probability can be estimated through random subsets)

Thank You for Your Attention!

Appendix

Literature

- H. Zou, “The adaptive lasso and its oracle properties,” Journal of the American statistical association, vol. 101, no. 476, pp. 1418–1429, 2006.
- H. Zou and H. H. Zhang, “On the adaptive elastic-net with a diverging number of parameters,” Annals of statistics, vol. 37, no. 4, p. 1733, 2009.
- C. S. Scheer, C. Fuchs, S.-O. Kuhn, M. Vollmer, S. Rehberg, S. Friesecke, P. Abel, V. Balau, C. Bandt, K. Meissner, et al., “Quality improvement initiative for severe sepsis and septic shock reduces 90-day mortality: a 7.5-year observational study,” Critical care medicine, vol. 45, no. 2, pp. 241–252, 2017.

Cox Model

Cox Model with its Proportional Hazard Assumption

$$h(t, X_i) = h_0(t)\exp(X_i\beta)$$

is not time dependent!

- h_0 as baseline hazard function
- $X_i = (x_{i1}, \ldots, x_{ip})$ covariates for subject i
- $\beta^T = (\beta_1, \ldots, \beta_p)$ coefficients vector $\leftarrow$ which has to be estimated

read: Healthcare Data Analytics, Reddy and Aggarwal, Chapman & Hall, 2015

Cox Model

Estimation of Regression Parameters

$$L(\beta) = \prod_{j=1}^{n} \left(\frac{\exp(X_j\beta)}{\sum_{i\in R_j} \exp(X_i\beta)} \right)^{\delta_i}$$

Observed event

Patients at risk

$$\ell(\beta) = \sum_{j=1}^{N} \delta_j \left(X_j\beta - \log \sum_{i\in R_j} \exp(x_i\beta) \right)$$

- δ_j is 0 for a censoring time, 1 otherwise
- R_j are the living individuals at time point j