Allianz

High-dimensional statistics and Machine learning with applications to Insurance

November 2022
Masaryk University, Brno

Ivana Milović, MAS PhD

Introducing Myself

Ivana Milović, MAS PhD

Non-Life Pricing Actuary (SME)
ivana.milovic@allianz.at

Prior experience

- Uniqa Insurance Group – Non-Life Pricing Actuary (Motor)
- Lecturer - University of Vienna
- Prae and Post-Doc Researcher - Department of Statistics,
  University of Vienna

Education

- PhD in Statistics (Univ. of Vienna, 2016)
- Master of Advanced Studies in Mathematics (Univ. of Cambridge, 2011)
- BSc in Mathematics and Computer Science (Univ. of Belgrade, 2010)

Allianz Group

Our Company

Allianz Group at a glance

With around **155,000** employees worldwide, the Allianz Group serves over **126** million customers[1] in more than **70** countries.

In fiscal year 2021 the Allianz Group achieved total revenues of approximately **148.5** billion euros.

Allianz is one of the world‘s largest asset managers, with third-party assets of nearly **2.0** trillion euros at year end.

[1]Including non-consolidated entities with Allianz customers
Data as of March 4, 2022 (release of the Annual Report 2021)

We are much more than just an insurer

… **Allianz** is the **Worldwide Insurance Partner of the Olympic & Paralympic Movements from 2021-2028** – and one of few global brands that can communicate with Olympic IP/rings

… **Allianz** insures major **Hollywood** and **Bollywood** movies, including all 24 James Bond productions

… **Allianz** supports sustainable motor sports by being a partner of the fully electric car racing Formula E Championship

… **Allianz** offers financial solutions to more than 49 million emerging consumers in Africa, Asia and Latin America

… **Allianz** is the leading specialist in **space insurance** and celebrated its **100th birthday as aviation insurer** in 2015

… **Allianz** insured the last three buildings to hold the **title of "world's tallest"**: Petronas Towers, Taipei 101 & Burj Khalifa

… the **Allianz Center for Technology** has conducted thousands of **crash test** since 1980 to improve **road safety**

… **Allianz** was one of the insurers of the **Titanic**

Allianz

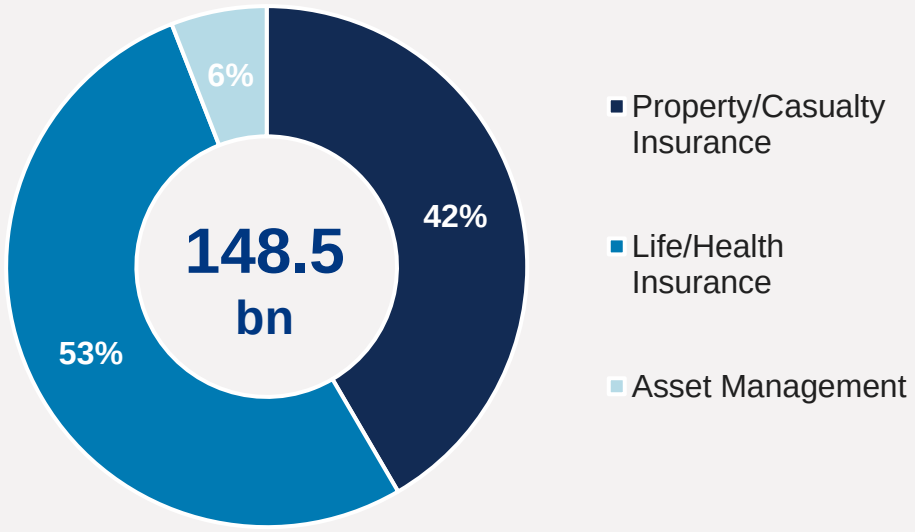
We are a global player

- Leading Property and Casualty insurer globally
- Among the top 5 Life/Health business globally
- Among the top 5 asset managers globally
- Global leader in credit insurance
- Worldwide leader in assistance services
- One of the leading corporate insurers globally

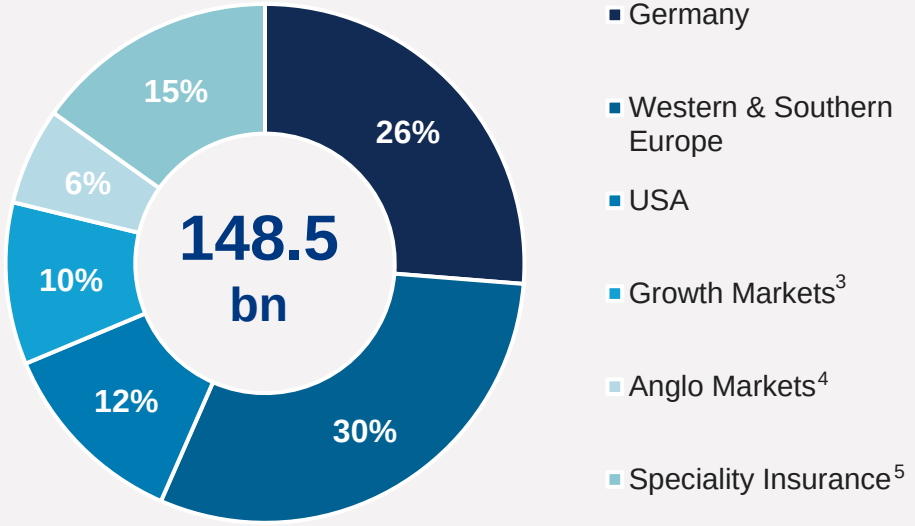
Revenues by segments and regions

Revenues Allianz Group 2021: EUR 148.5 bn

by SEGMENTS



by REGIONS[1,2]



[1] Excl. Corporate & Other and consolidations
[2] Incl. Banking

[3] Central and Eastern Europe, Asia-Pacific, Latin America, Middle East and Africa, Turkey. Austria and Allianz Direct allocated to Western and Southern Europe.
[4] UK, Ireland, Australia
[5] Allianz Global Corporate & Specialty, Euler Hermes, Allianz Partners, Allianz Re

Data as of March 29, 2022 (release of the Allianz Fact Sheet; further information on allianz.com)

Interested in joining one of our companies?

- ❑ Bachelor/Master Thesis supervision
- ❑ Summer internship
- ❑ Full- or part time job

Contact us:

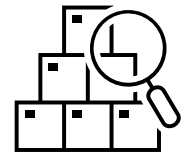
ivana.milovic@allianz.at

What is pricing?

Allianz

What is pricing?

“Pricing is the way that a company decides prices for its products or services, or the prices decided” – Cambridge dictionary



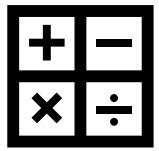
€

Why do we need statistics and mathematical modelling for pricing in insurance?

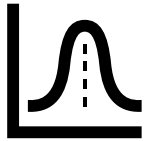
Because the cost of a policy is random. How do we estimate it?

There are two ways:

- Based on the historical data/expert judgement (simplistic approach)



- Fitting statistical models to historical data -> technical pricing.



Summary from yesterday

Allianz

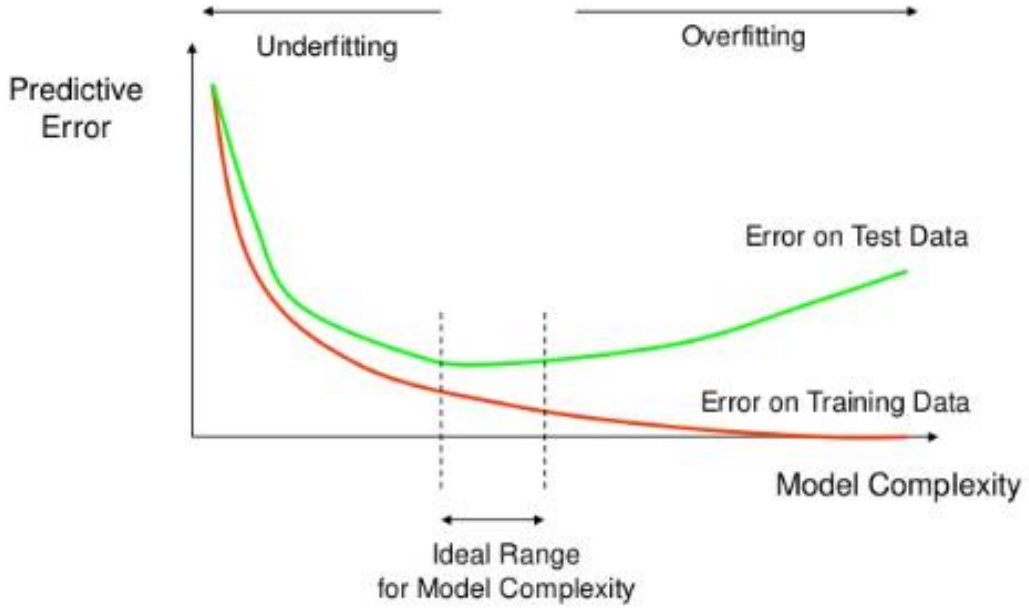
Summary

- We assess the model quality by its **prediction error**

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}(X_i))^2$$

  given a sample $(X_i, Y_i)_{i=1}^{n}$.
- But this is only one part of it – **training (in-sample) error**
- It is necessary to estimate this error for new (unseen) data – **testing (out-of-sample) error**

Summary



A model (and its complexity) should be chosen based on these two prediction errors:

Allianz

Summary

- The training error we can estimate from the sample directly
- There are two types of methods for estimating the testing error
  1. Cross – validation: based on **resampling**
  2. AIC, BIC, etc.: based on **testing error ≈ training error + dimension penalty**

Content

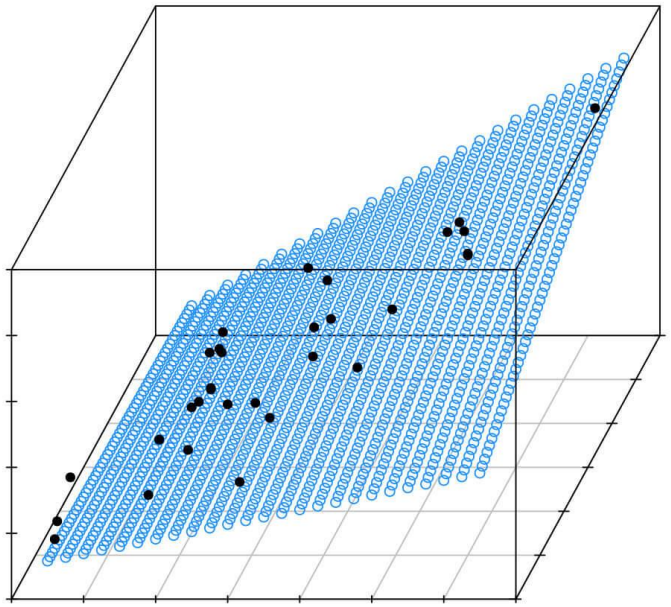
Topics

- ~~Model assessment and selection~~
- ~~Cross validation, AIC, BIC~~
- Linear Models
- PCR, Regularization methods
- Generalized Linear models
- Pricing process
- Machine Learning in Insurance

Types of Models

Allianz

Linear Models



Allianz

Model selection and regularization

- **Linear models** (and generalized linear models: GLMs), though simple, turn out to be surprisingly competitive in real-world problems, compare to more complex models (GLMs are the standard in the insurance business)

- Reason for that lies in their simplicity and interpretability

- But what is their prediction accuracy and what happens when the number of parameters $\boldsymbol{p}$ **is large compared to the sample size** $\boldsymbol{n}$?

Model selection and regularization

Allianz

- Let us focus on linear models, for demonstration
- Assume that: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, for some $\beta \in R^p$

  $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma I$.

  Also, $\mathrm{Y} \in R^n$ and $\mathrm{X} \in R^{n \times p}$.

- If $n \geq p$ then the OLS estimator $\hat{\beta} = (X'X)^{-1}X'Y$ is well-defined and it is unbiased.
- Therefore, the estimates $\hat{Y} = X\hat{\beta}$ are unbiased.
- But for $p > n,$ OLS is not even defined. Therefore, we have to come up with some other estimators.

Allianz

Model selection and regularization

But what about the **variance** of these estimates?

- If $n \gg p,$ the variance is usually small, and our estimates are accurate

- But if two or more variables are **highly correlated**, this could lead to high variance and therefore unstable estimates. This happens, because $\det(X'X)$ is almost 0 and the matrix inversion becomes very unstable

Model selection and regularization

Example of (potentially) **highly-correlated variables** in Motor Insurance

- Vehicle age and contract age
- Engine Power and engine volume
- Population density and regional segmentation variables

Example of (potentially) **highly-correlated variables** in SME Insurance

- Turnover and number of employees

Allianz

Model selection and regularization

- Also, if n is not much larger than p, the estimates can get very unstable.
- Example: if all regressors are i.i.d. N(0,1) the variance of the predictions equals $\sigma \frac{p}{n-p-1}$.
- This is problematic for p large compared to n.

Allianz

Model selection and regularization

- To summarize, we need to find methods to

  - ❑ reduce the number of parameters (dimensionality) and/or
  - ❑ to reduce the variance of the estimators

- Otherwise, the models are unreliable!

Allianz

Model selection and regularization

Alternatives to OLS in linear regression:

- Subset selection (best subset and stepwise)
  - The goal is to choose a subset of all regressors and that that model as approximation
  - Cross Validation, AIC or BIC help us then choose the best submodel
- Dimension reduction (PCA, for example)
  - We transform the original regressors so that the new ones are uncorrelated and sorted in order of importance, so we can also reduce the dimension
- Shrinkage methods (Ridge, Lasso, etc.)
  - Modification of OLS (Ordinary Least Squares) estimator, so that the coefficients of the estimates are shrunk towards zero, to reduce the variability of the estimator and make them more stable
  - Also works for p larger than n

Allianz

Model selection and regularization

For more details, see the Appendix chapter

GLM – industry standard

Allianz

GLM

- Generalized linear models (GLM) are a natural extension of linear models
- Response variable is now **function** of a **linear combination** of regressors
- Response variable does not have to be distributed normally anymore, it can take on of the distributions from the **exponential family**: Bernoulli, Binomial, Poisson, Gamma, Exponential
- GLMs are widely used in insurance industry and are ideally suited for the analysis of the non-normal data, that is commonly encountered in insurance.

GLM - Model choice

Allianz

$\underline{Y}$	Claim frequencies	Claim numbers or counts	Average claim amounts	Probability (eg of renewing)
Link function $g(x)$	ln(x)	ln(x)	ln(x)	ln(x/(1-x))
Error	Poisson	Poisson	Gamma	Binomial
Scale parameter ϕ	1	1	Estimated	1
Variance function $V(x)$	x	x	x^2	x(1-x)*
Prior weights $\underline{\omega}$	Exposure	1	# of claims	1
Offset $\underline{\xi}$	0	ln(exposure)	0	0

* where the number of trials=1, or x(t-x)/t where the number of trials = t

Allianz

GLM

Generalized Linear Models serve as the industry standard for non-life insurance pricing

- Their multiplicative output remains understandable also for non-actuaries
- Range of professional insurance software dedicated to GLM
- GLM is also possible in R, Python and other open-source software

Allianz

Risk Premium

- In order to come up with a price for an insurance policy, we need to start by modelling the base risk premium, i.e.

- **(Net) Risk Premium** is defined as **Frequency (how often)**× **Severity (how high)**

- One can model
  - ❑ the frequency of claims -> Poisson
  - ❑ the claim amount (severity)-> Gamma
  - ❑ (directly) the risk premium -> Gamma, Tweedie

Risk Premium

- This procedure has to be done for every insurance coverage.
- Example in SME insurance

	COMFORT My Company	PLUS My Company	EXTRA My Company	MAX My Company
Business liability	✓	✓	✓	✓
Fire	✓	✓	✓	✓
Storm and extraordinary natural events	✓	✓	✓	✓
Basic Assistance	✓	✓	✓	✓
Water damage		✓	✓	✓
Glass breakage		✓	✓	✓
Burglary and robbery			✓	✓
Malicious damage			✓	✓
Technical dangers				✓
IT-Assistance				✓

Allianz

Risk Premium

Data preparation

Initial Analysis

GLM possible ?

GLM analysis

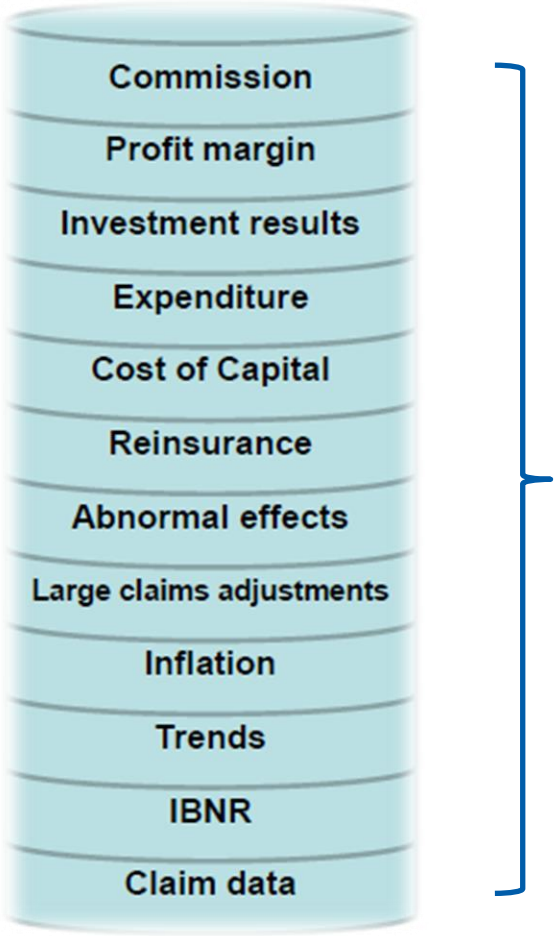
Simplified Pricing Method

Net risk premium

Risk Premium

- In the end we need to deliver a final risk premium
- We should combine all models we made
- Necessary to understand the total effect
- Result: total net risk premium

Risk Premium: from net to gross



A whole range of effects is to be added to the net risk premium

Gross Risk Premium

GLM – space for improvement

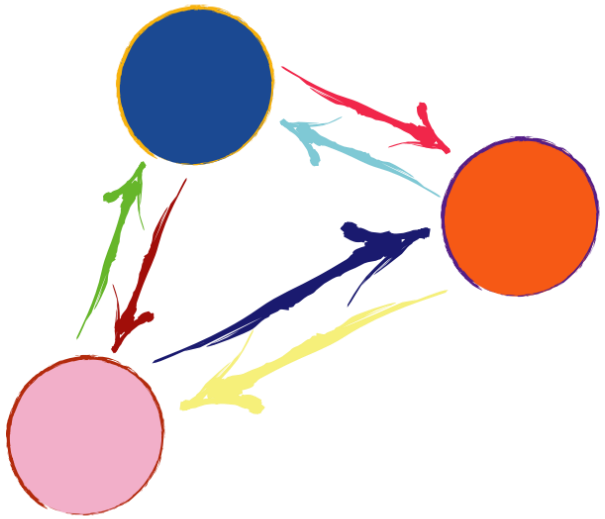
Allianz

Interactions and GLM

- An **interaction** effect exists when the effect of an independent variable on a dependent variable changes, depending on the value(s) of one or more other independent variables

- In that case an interaction term(s) has to be added to the model

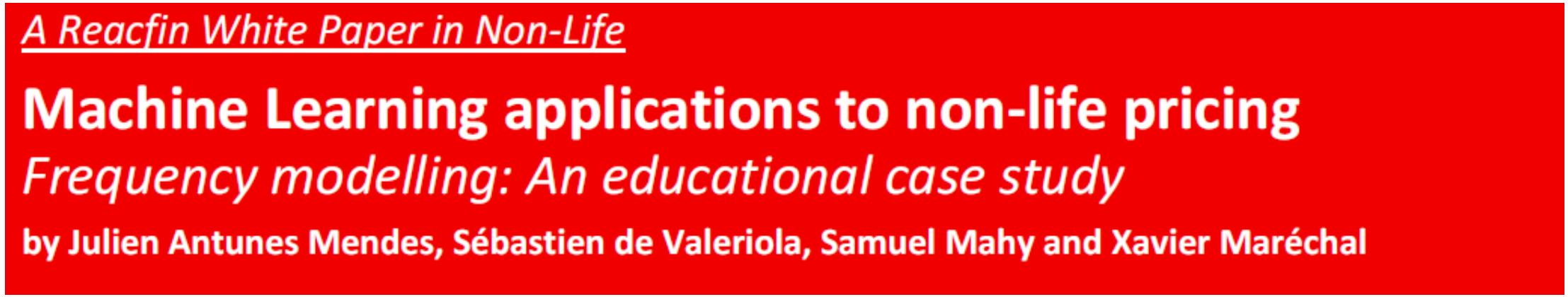
- Example: gene A and gene B may contribute to developing a certain disease, but in combination they are fatal



Allianz

Interactions and GLM

- The problem? GLM models do not detect interactions automatically
- Then can be added to the model, but this has to be done ‘manually’
- Example taken from:



Interactions and GLM

The simplified frequency database we used was sampled using a Poisson distribution function where the Poisson frequency parameter λ was designed as a function of two explanatory variables: Age and Power. We simulated ages and powers, and then computed the following frequencies:

$$\lambda = a\,(age - b)^2 + c\,power + d\,I_{\{age \geq 60\} \cap \{power \geq 50\}},$$

where a, b, c, d are positive real parameters calibrated in such a way that the range of λ is consistent with a frequency range

As shown in Figure 3, the Poisson frequency surface includes a non-linear interaction between the two explanatory variables and has been chosen on purpose to « fail » standard statistical methods (as GLM) and therefore show how some machine learning methods can « fix » these issues.

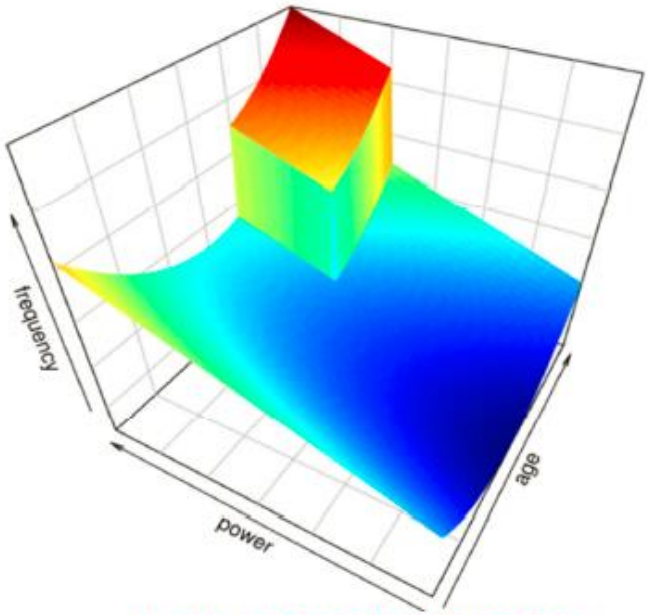


Figure 3 : Poisson frequency surface

Allianz

Interactions and GLM

- In this example: there is an interaction of age and engine power

$$\boldsymbol{Age \geq 60 \; and \; Engine \; Power \geq 50}$$

- But if this effect is not noticed and included in the model, the GLM fit is poor

Interactions and GLM

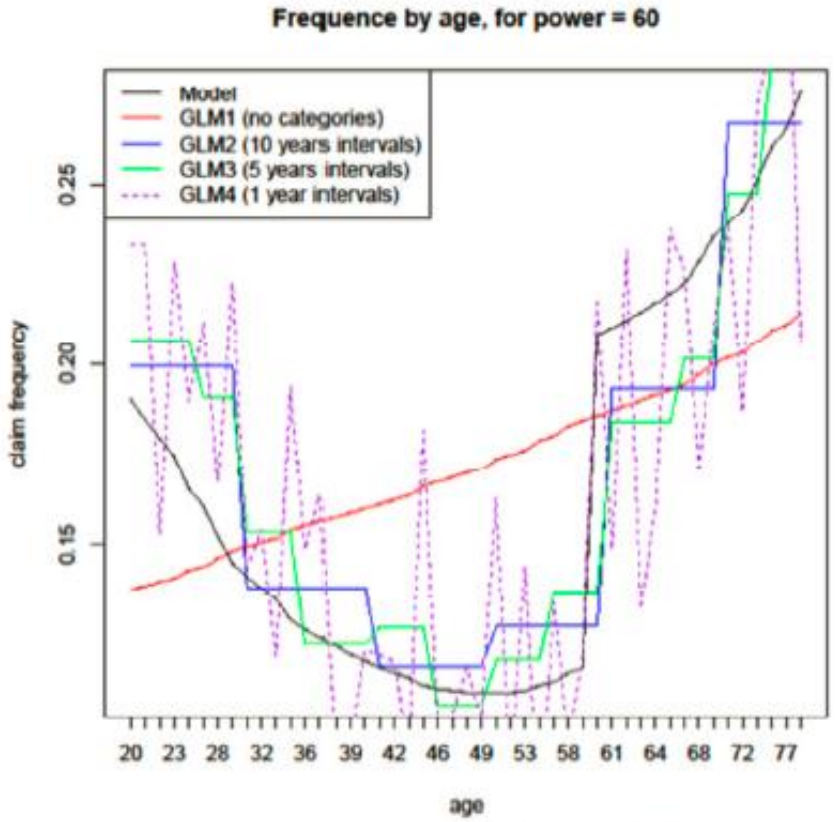


Figure 8: Frequency by age with GLM

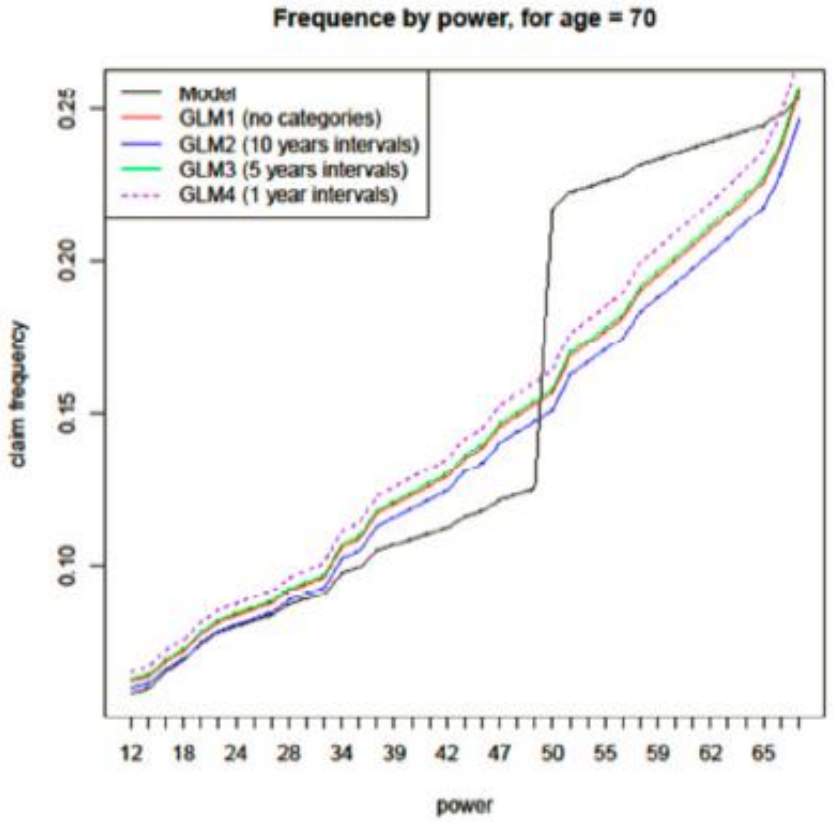


Figure 9: Frequency by power with GLM

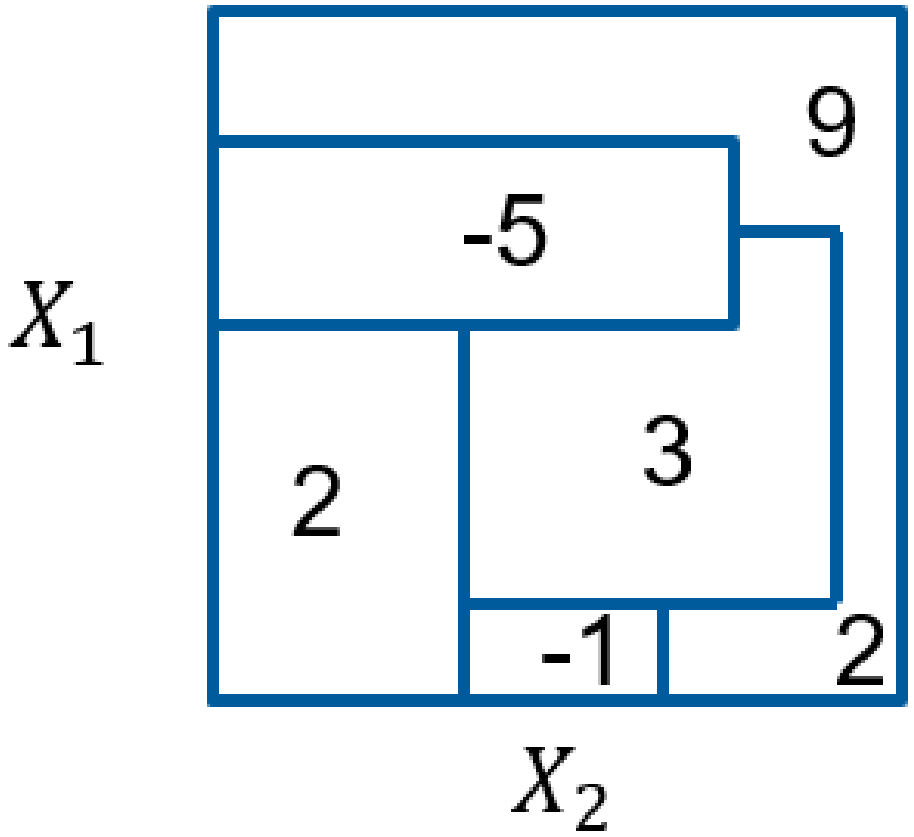
Tree based methods

- But many machine learning algorithms can automatically capture these effects
- Let us take Gradient Boosting Trees for example
- How does this algorithm work?
- Let us present some basics

Allianz

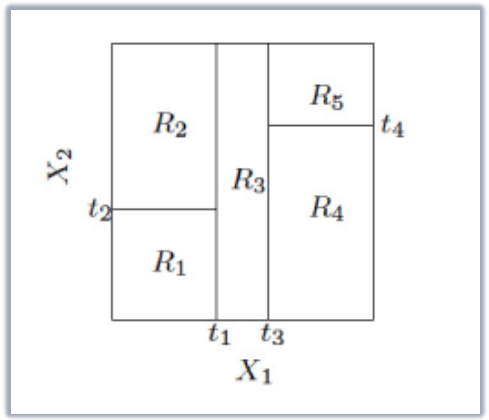
Tree based methods

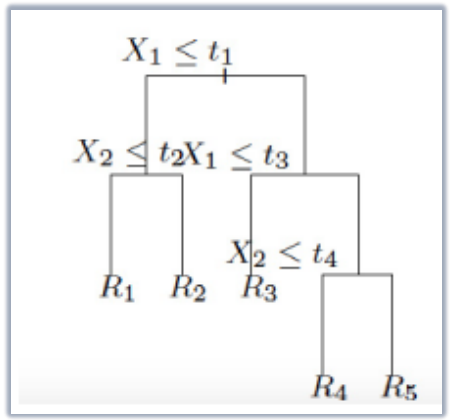
- Tree-based methods partition the feature space into a set of rectangles and then fit a simple model (typically a constant) in each region
- Consider a regression problem with continuous response Y and continuous regressors $X_1, X_2 \in (0,1)$.
- For example, this partition is simple, but cannot be obtained by recursive binary splitting, i.e. represented by a tree.



Tree based methods

- So, let us restrict our attention to **recursive binary partitions**, like this one:

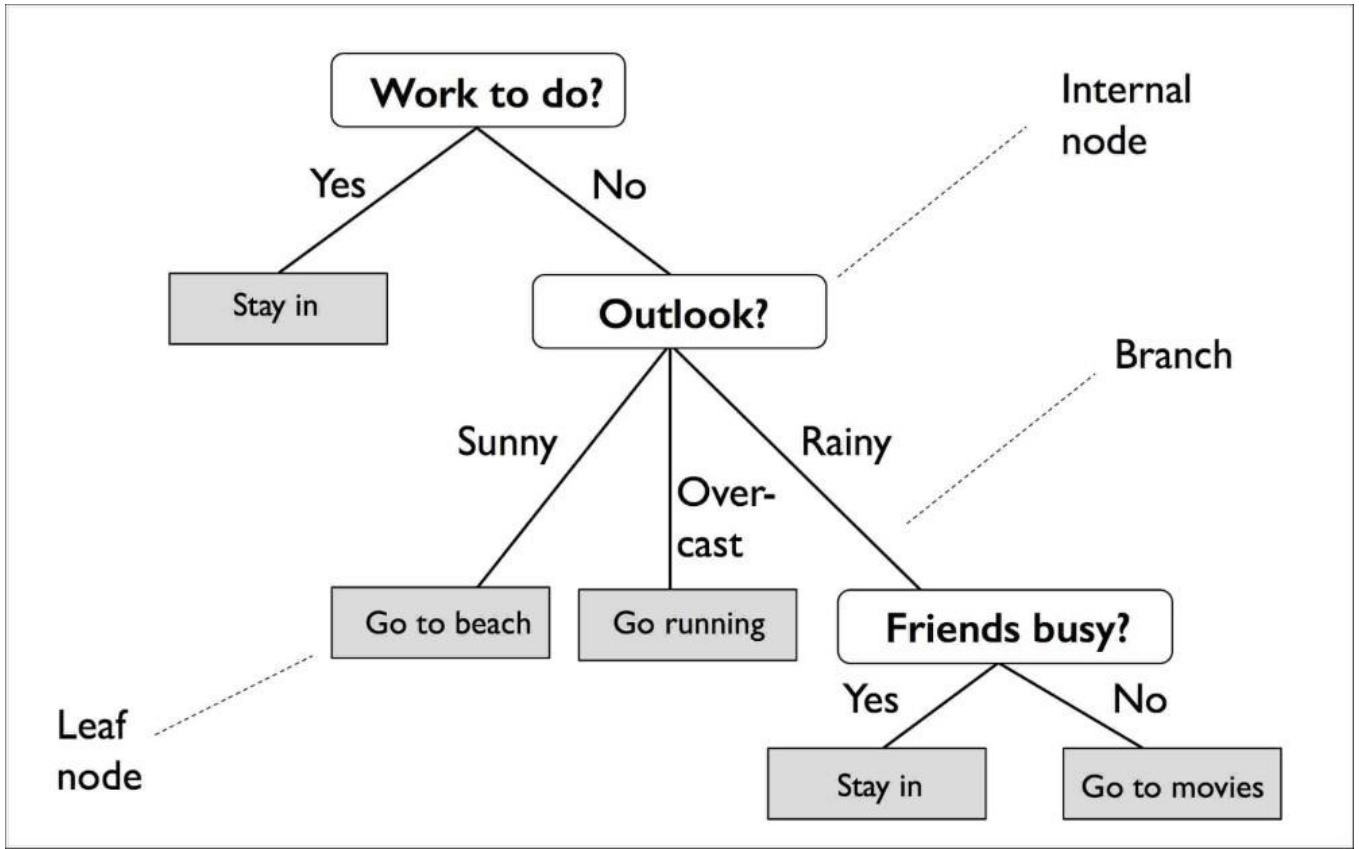




- First split the space into two regions and model the response by the mean of Y in each region. Choose the split variable and split-point to achieve the optimal split.
- Then one or both regions are further split in the same fashion iteratively until some stopping rule is applied.

Tree based methods

- The corresponding regression model predicts Y with a constant c_m if the inputs X are in region R_m, i.e.



Tree based methods

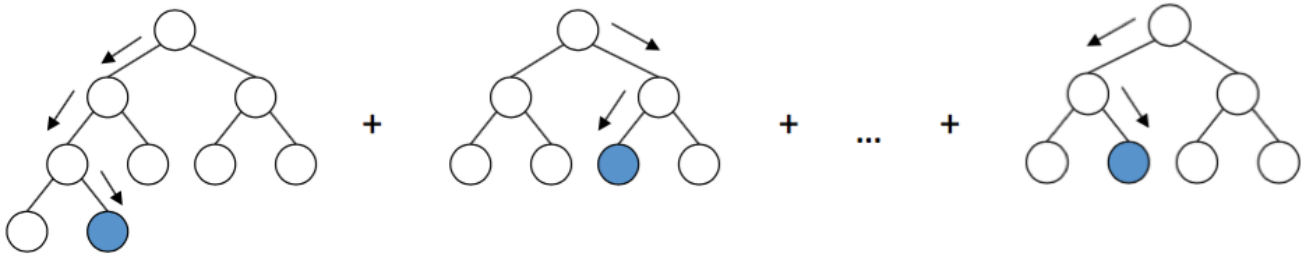
These trees can now further be used for boosting

What is boosting?

Allianz

Boosting

- Gradient boosting is one of the most powerful techniques for building predictive models. It is proven successful in many areas and is one of the leading methods for winning Kaggle competitions (https://kaggle.com/)



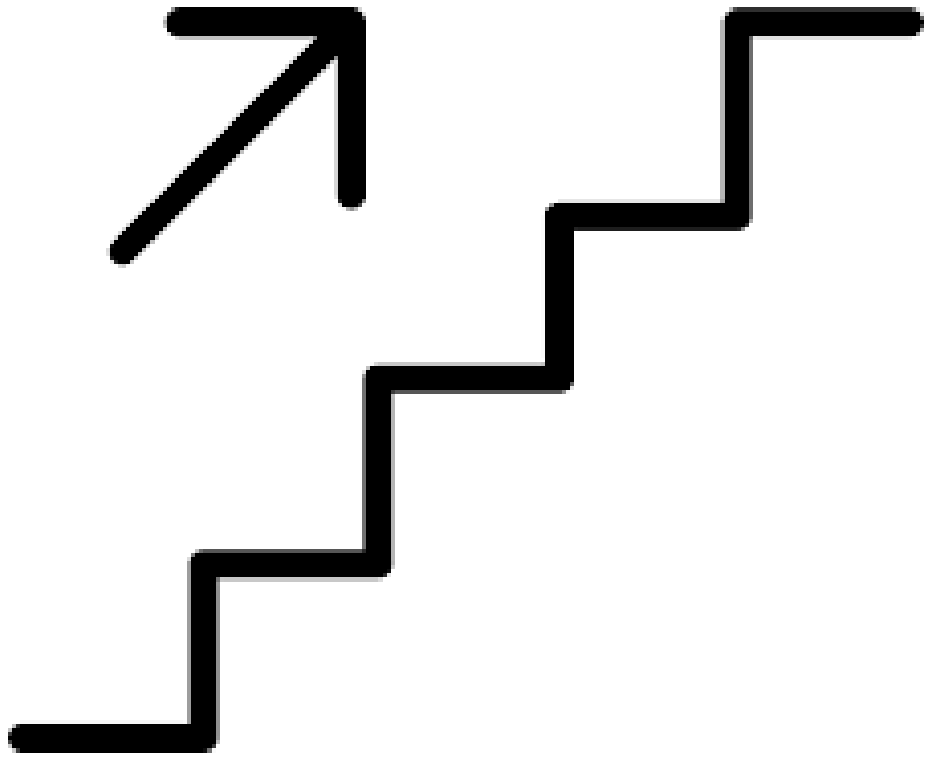
Allianz

Boosting

- In general: models can be fitted to data individually or combined in an ensemble – a combination of simple individual models (usually trees) that together create a more powerful model
- Boosting is a method that builds the model in a **stage-wise fashion**.
- It starts by fitting an initial model.
- The second model focuses then on accurately predicting the cases where the first model performed badly
- The third model focuses on correcting the faults of the previous stage, etc.

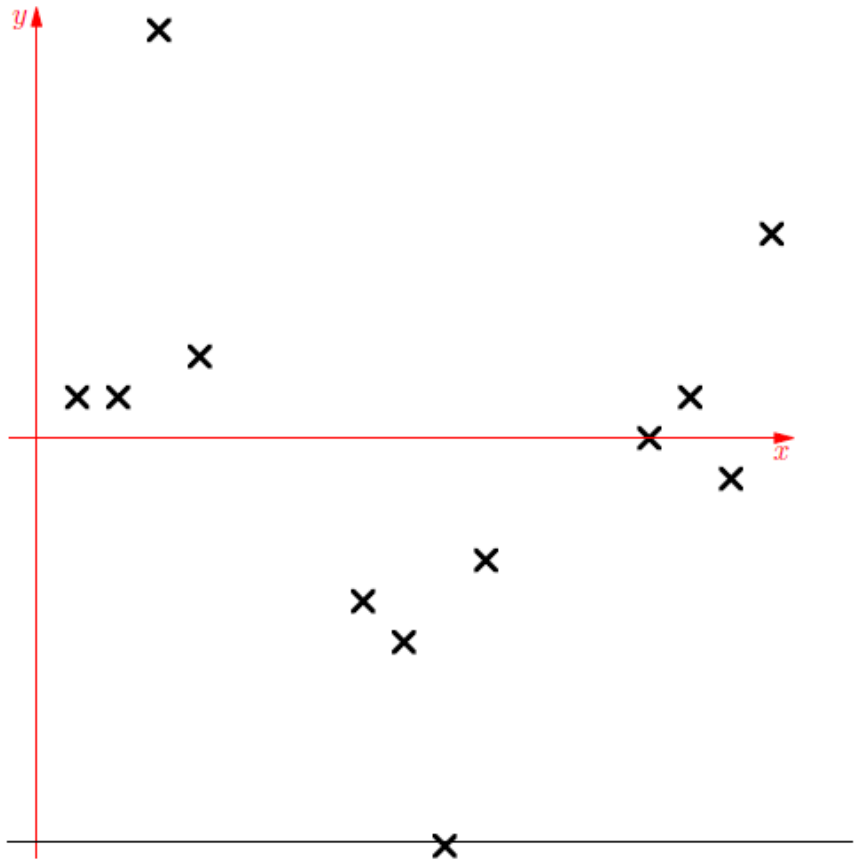
Boosting

- Here we do not fit one big decision tree to the model, because this can easily lead to overfitting
- Instead, the boosting algorithm learns slowly
- At each step we fit a decision tree to the residuals from the previous model
- Then new tree is then added to the model



Boosting

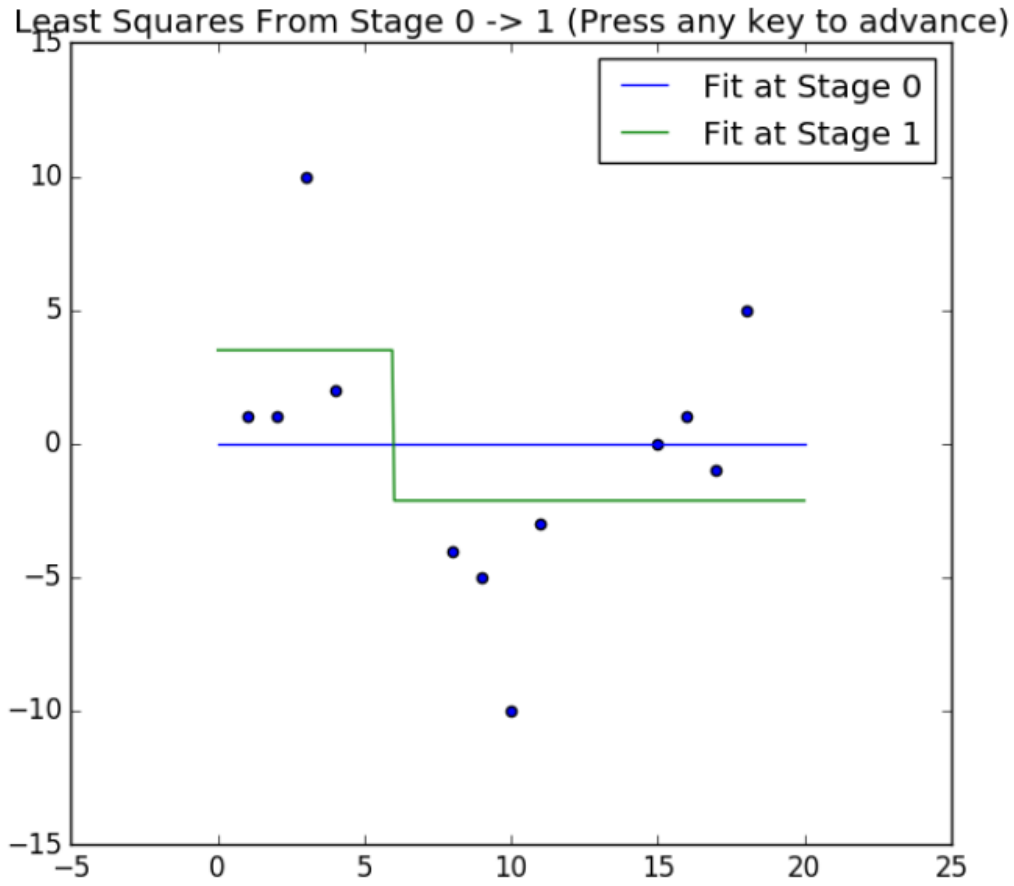
Example: data to be fitted



Plot courtesy of Brett Bernstein.

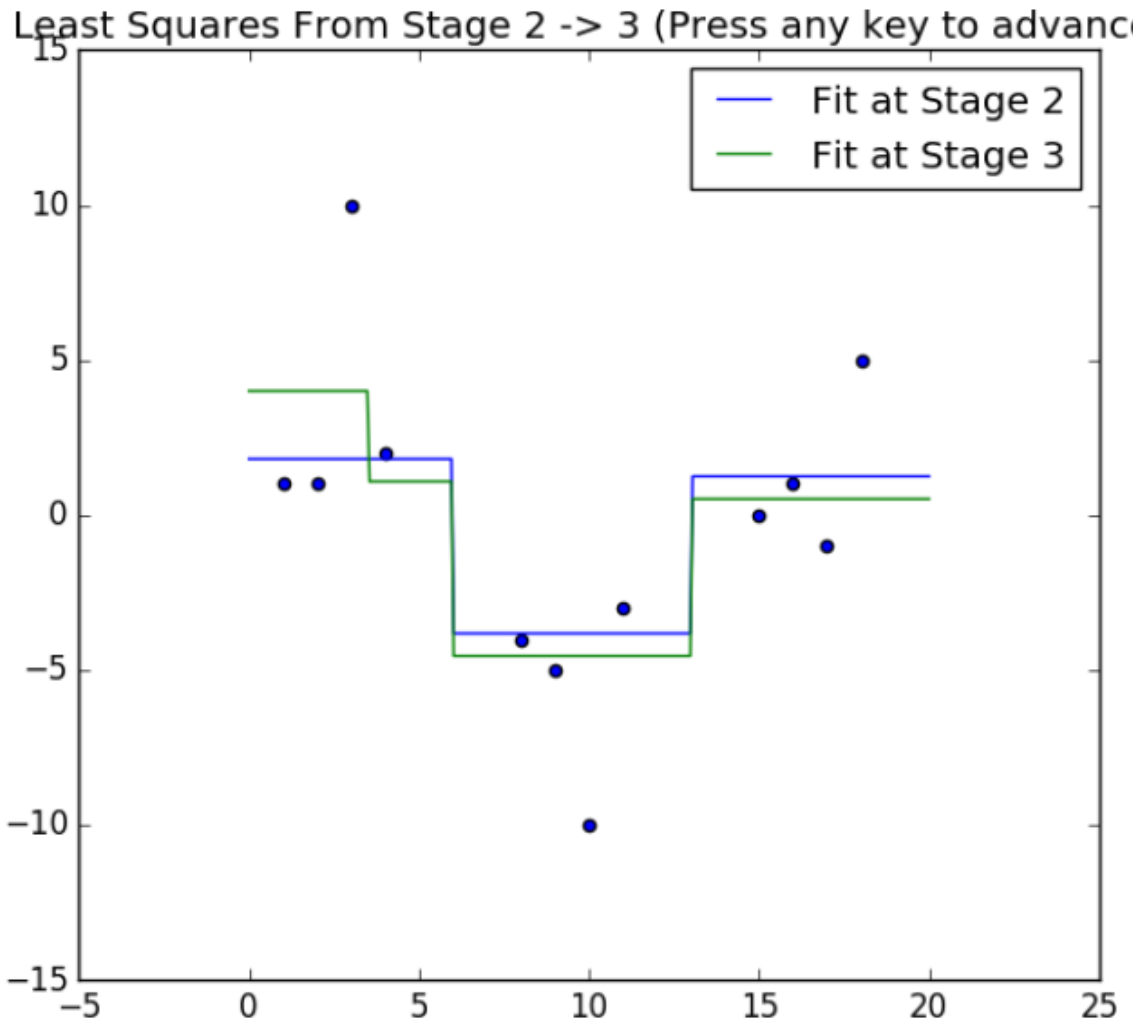
Allianz

Boosting



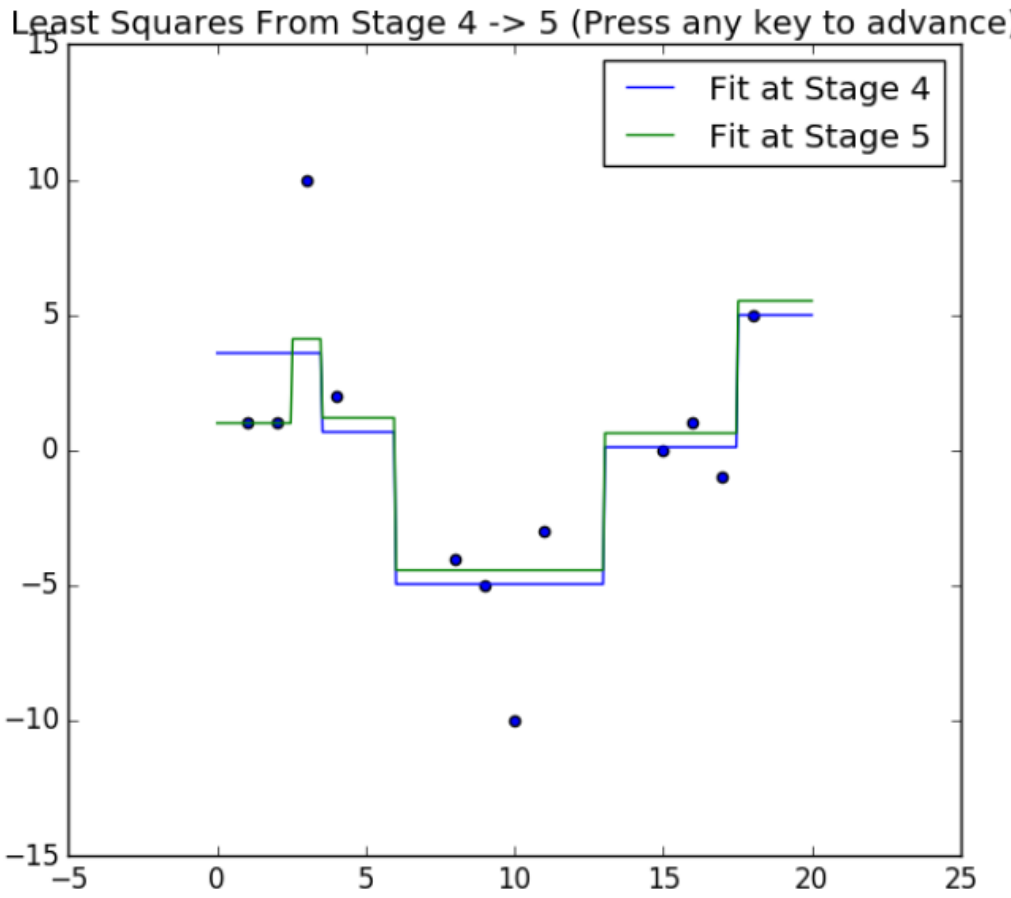
Allianz

Boosting



Allianz

Boosting



Allianz

Boosting

Usually the trees are rather small, but they should be deep enough to capture interactions. Number of splits = 2 is already enough to catch first-order interactions

There are several parameters that need to be chosen: the number of trees, the number of splits in each tree and the learning rate of the algorithm (usually 0.1 or 0.01)

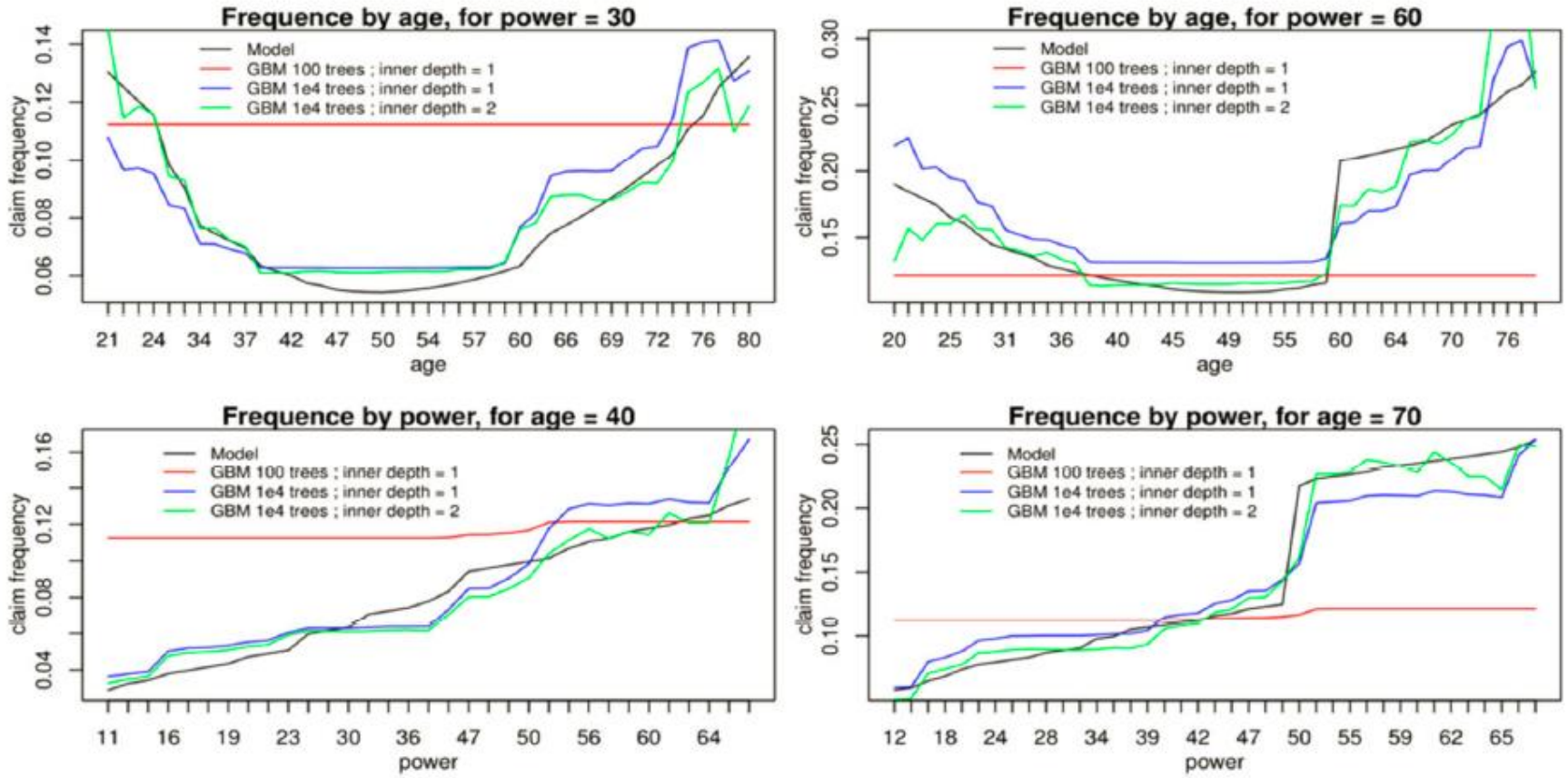
For the number of trees **cross-validation** is used

Example

- Back to our example
- Remember that GLM could not ‘recognize’ the interaction between age and engine power
- But GBMs do, provided that the tuning parameter have be carefully selected

Example

SUCCESS!



Allianz

GLM vs Machine Learning

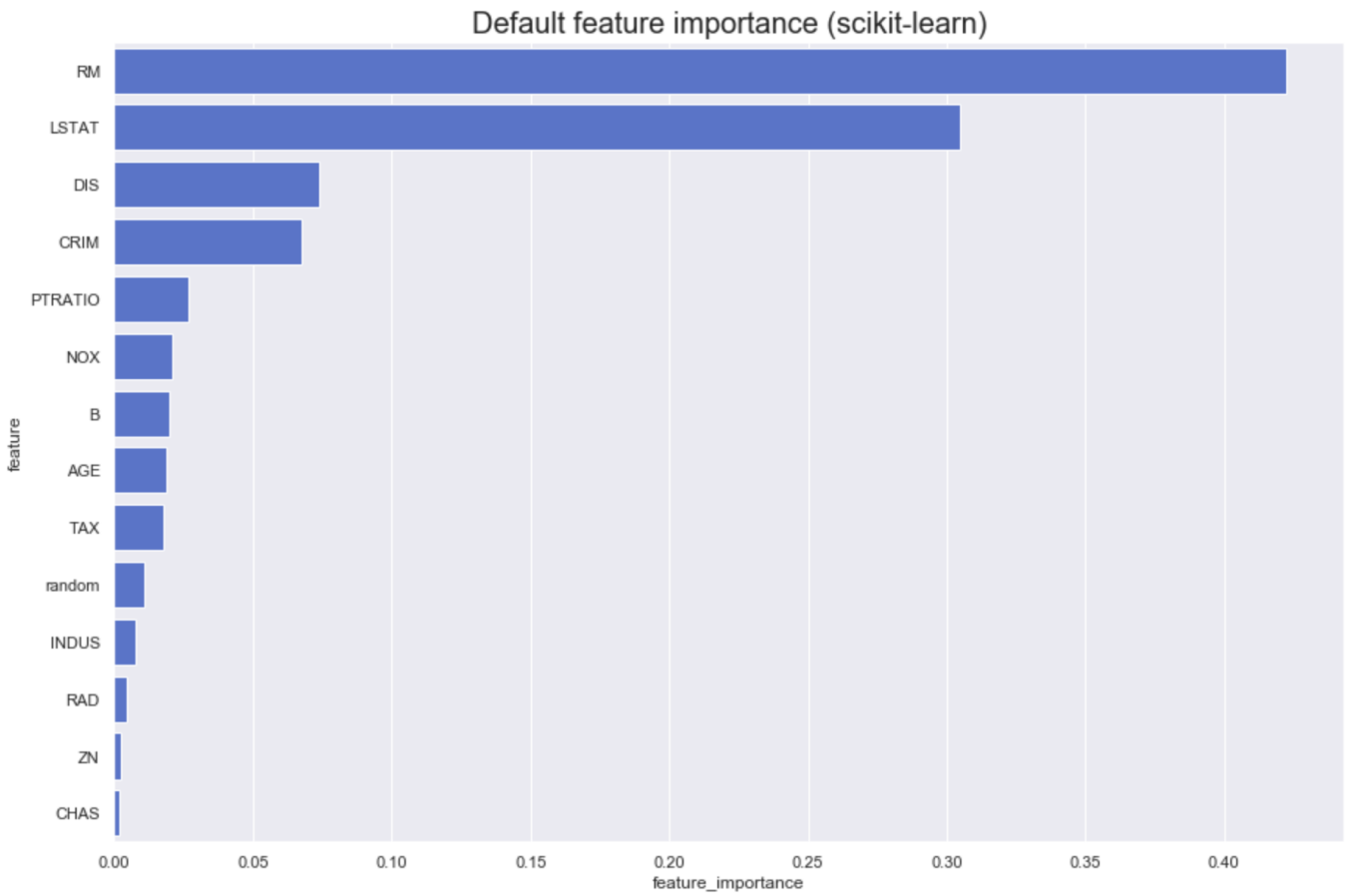
- The problem with these kind of algorithms is that the interpretation is almost completely lost
- It is very unlikely that such models will be approved by regulators, at least in the majority of countries
- And even if they are, then the insurance company runs into the risk of reputational loss, in case some of the ethical problems discussed before emerge

Allianz

GLM vs Machine Learning

- Also, the actuaries want to understand their models and not use black-box alternatives
- So, GLMs will probably not be replaced by Machine Learning algorithms in the near future
- But they can assist the actuaries in spotting interactions, as well as determine variable significance or perform clustering tasks

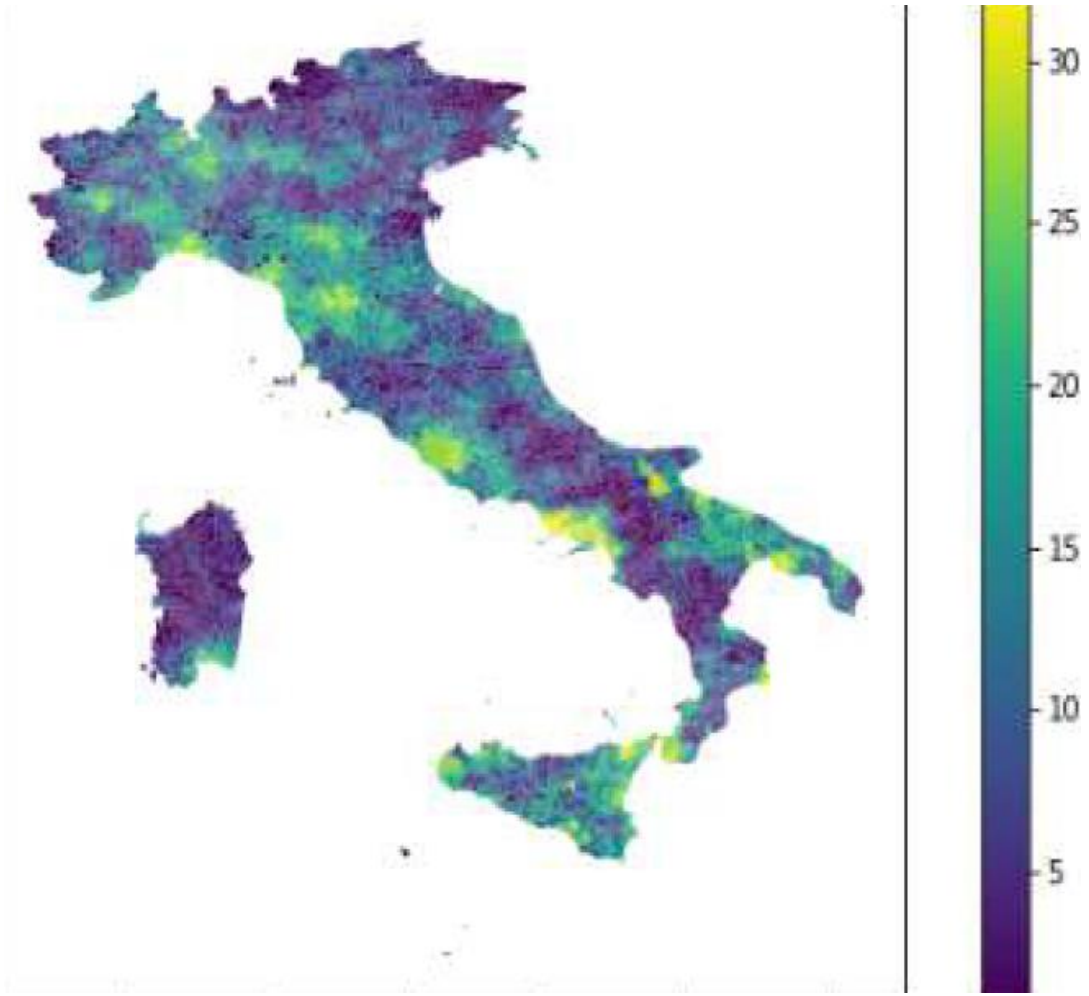
Variable importance



Clustering

- Examples of clustering can be brand, region or business activities clustering.

- Here the black-box nature of the models is not so important, because the model results can usually easily be validated



Literature

- An Introduction to Statistical Learning with Applications in R - Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
- https://www.reacfin.com/wp-content/uploads/2016/12/20170914-Machine-Learning-applications-for-non-life-pricing.pdf
- https://www.stat.cmu.edu/~ryantibs/datamining/lectures/17-modr2.pdf

Appendix

Subset Selection

Subset Selection

Allianz

1. **Best subset selection**: for a linear model with p predictors do

- Let M_0 be the null model with zero regressors, i.e. sample mean of Y is used as a predictor
- For $k = 1,2, \dots, p$
  1. Fit all $\binom{p}{k}$ models that contain exactly k predictors
  2. Pick the best among these $\binom{p}{k}$ models and call it M_k. I.e., choose the model with the largest R^2.
  - Select the best model from $M_0, M_1, \dots, M_p$ using cross-validation, AIC, BIC, etc.
  - Note: here you cannot use R^2 because then the largest model would always be chosen.

https://en.wikipedia.org/wiki/Coefficient_of_determination

Allianz

Subset Selection

- This method is conceptually very simple to understand
- Problem? Too many models to fit!
- How many? 2^p models to fit.
- For example: for $p = 40$, there are 1 073 741 824 models to fit!
- So, we need another solution.

Allianz

Subset Selection

2. Stepwise selection
   - ➢ Forward
   - ➢ Backward

Forwards stepwise selection

- Computationally efficient alternative to the best subset selection
- Here we begin with the null model and add predictors **one at the time** until we get the full model (or some stopping rule is applied)
- Then we choose among these models using cross-validation, AIC, BIC, etc.

Allianz

Subset Selection

More formally:

Forwards stepwise selection: for a linear model with p predictors do

- Let M_0 be the null model with zero regressors, i.e. sample mean of Y is used as a predictor
- For $k = 0,1, \dots, p-1$
  1. Consider all $p-k$ models that add one additional predictor to the model M_k
  2. Pick the best among these $p-k$ models and call it M_{k+1}. I.e. choose the model with the largest R^2.
  - Select the best model from $M_0, M_1, \dots, M_p$ using cross-validation, AIC, BIC, etc.
  - Note: here you cannot use R^2 because then the largest model would always be chosen.

Allianz

Subset Selection

- Here we fit only $1 + \sum_{k=0}^{p-1}(p-k) = 1 + \frac{p\,(p+1)}{2}$ models
- For example: for $p = 40$, there are **466** models to fit. Much better than before.
- This procedure works well in practice, but now there is no guarantee that we will select the best method overall

Backwards stepwise selection:

Similar: here you start with the full model and delete regressors one at the time

Allianz

Example: Prostate cancer

- The data come from a study that examined the correlation between the level of prostate specific antigen (response variable) and a number of clinical measures (regressors) in men who were about to receive a radical prostatectomy.
- It is data frame with 97 rows and 9 columns.

Example: Prostate cancer

Allianz

These data come from a study that examined the correlation between the level of prostate specific antigen and a number of clinical measures in men who were about to receive a radical prostatectomy. It is data frame with 97 rows and 9 columns.

Usage

```
data(Prostate)
```

Format

The data frame has the following components:

`lcavol`
: log(cancer volume)

`lweight`
: log(prostate weight)

`age`
: age

`lbph`
: log(benign prostatic hyperplasia amount)

`svi`
: seminal vesicle invasion

`lcp`
: log(capsular penetration)

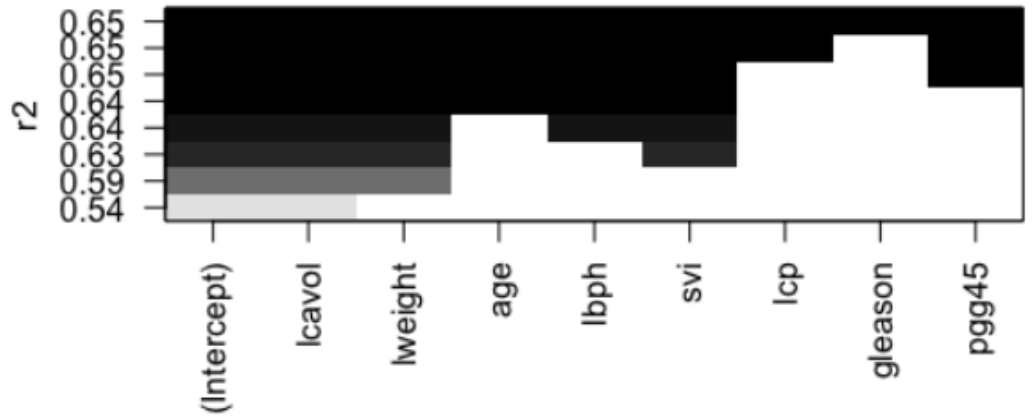
`gleason`
: Gleason score

`pgg45`
: percentage Gleason scores 4 or 5

`lpsa`
: log(prostate specific antigen)

Example: Prostate cancer

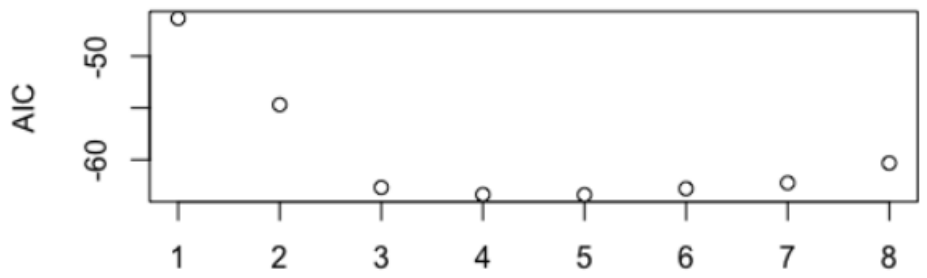
R Package **Leaps** is used to select the best model (based on R^2) of each size

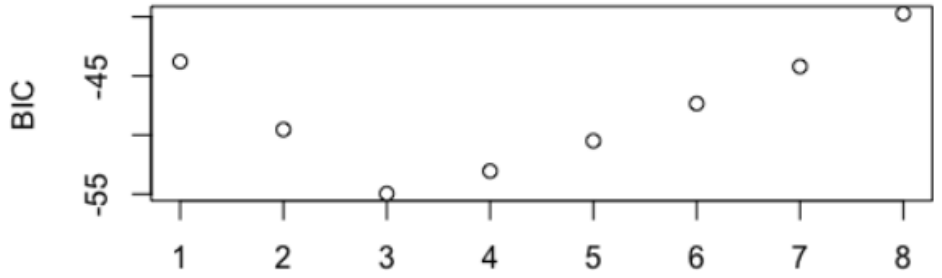


Allianz

Example: Prostate cancer

- Then AIC and BIC are calculated for each of these models, based on the formula for linear regression with normal errors.






```
> v<-leaps.out$which[which.min(AIC),] #which variables are chosen T/F
> names(X)[v] #gives us the names of those variables
[1] "lcavol"  "lweight" "age"     "lbph"    "svi"
> v<-leaps.out$which[which.min(BIC),] #which variables are chosen T/F
> names(X)[v] #gives us the names of those variables
[1] "lcavol"  "lweight" "svi"
```

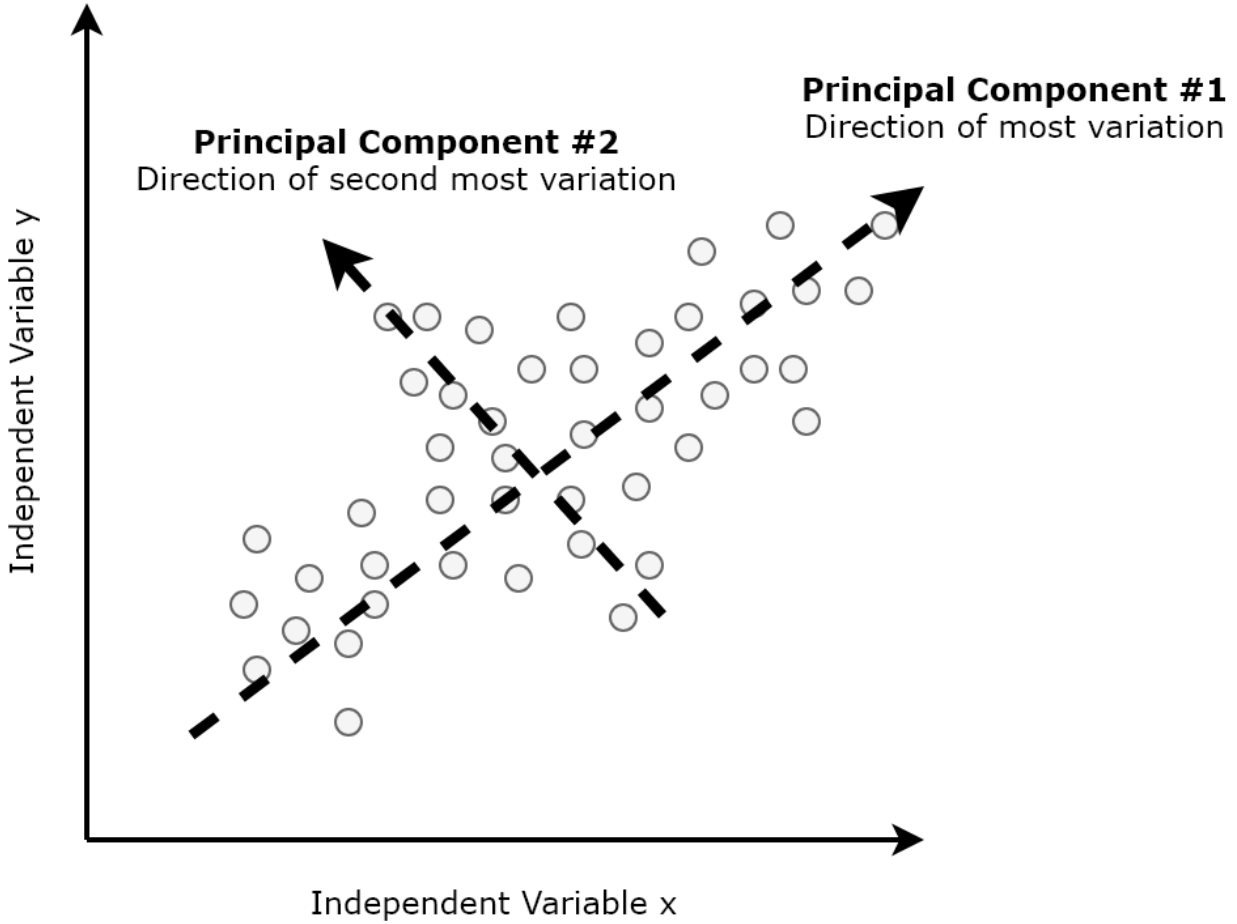
Preview

We are still to see:

- Some other methods that do model selection for linear models
- How to deal with correlations
- How to deal with $p > n$ case?

PCA

Principal Component Regression



Allianz

Principal Component Regression

- PCA uses an orthogonal transformation to convert a set of possibly correlated variables into a set linearly uncorrelated variables called **principal components**.

- This transformation is defined in such a way that the first principal component has the largest variance, the second principal component the second largest, etc.

Principal Component Regression

Allianz

- This way a dimension reduction can be performed and consequently OLS can be fitted using the newly obtained regressors.
- One can show that this reduces the variance of the OLS estimator
- But, interpretability issues!

Shrinkage Methods

Shrinkage Methods

Allianz

- We have already mentioned that if p is relatively large compared to n, or if some regressors are highly-correlated then the OLS estimates can be very variable and therefore unstable.
- Also, we cannot do OLS for $p > n$.
- In order to tackle these problems, shrinking the regression coefficients is helpful

Shrinkage Methods

Allianz

- **Idea: introduce some bias, but decrease variance significantly**
- This is done by adjusting our minimization problem

Shrinkage Methods

Allianz

- OLS:

$$\hat{\beta} = \underset{\beta}{argmin} \, \|Y - X\beta\|_2^2$$

- Ridge:

$$\hat{\beta}(\lambda) = \arg\min_{\beta}(\underbrace{\|Y - X\beta\|_2^2}_{\text{goodness of fit}} + \underbrace{\lambda \|\beta\|_2^2}_{\text{penalty}})$$

- Lasso:

$$\hat{\beta}_{lasso}(\lambda) = \arg\min_{\beta}(\underbrace{\|Y - X\beta\|_2^2}_{\text{goodness of fit}} + \underbrace{\lambda \|\beta\|_1}_{\text{penalty}})$$

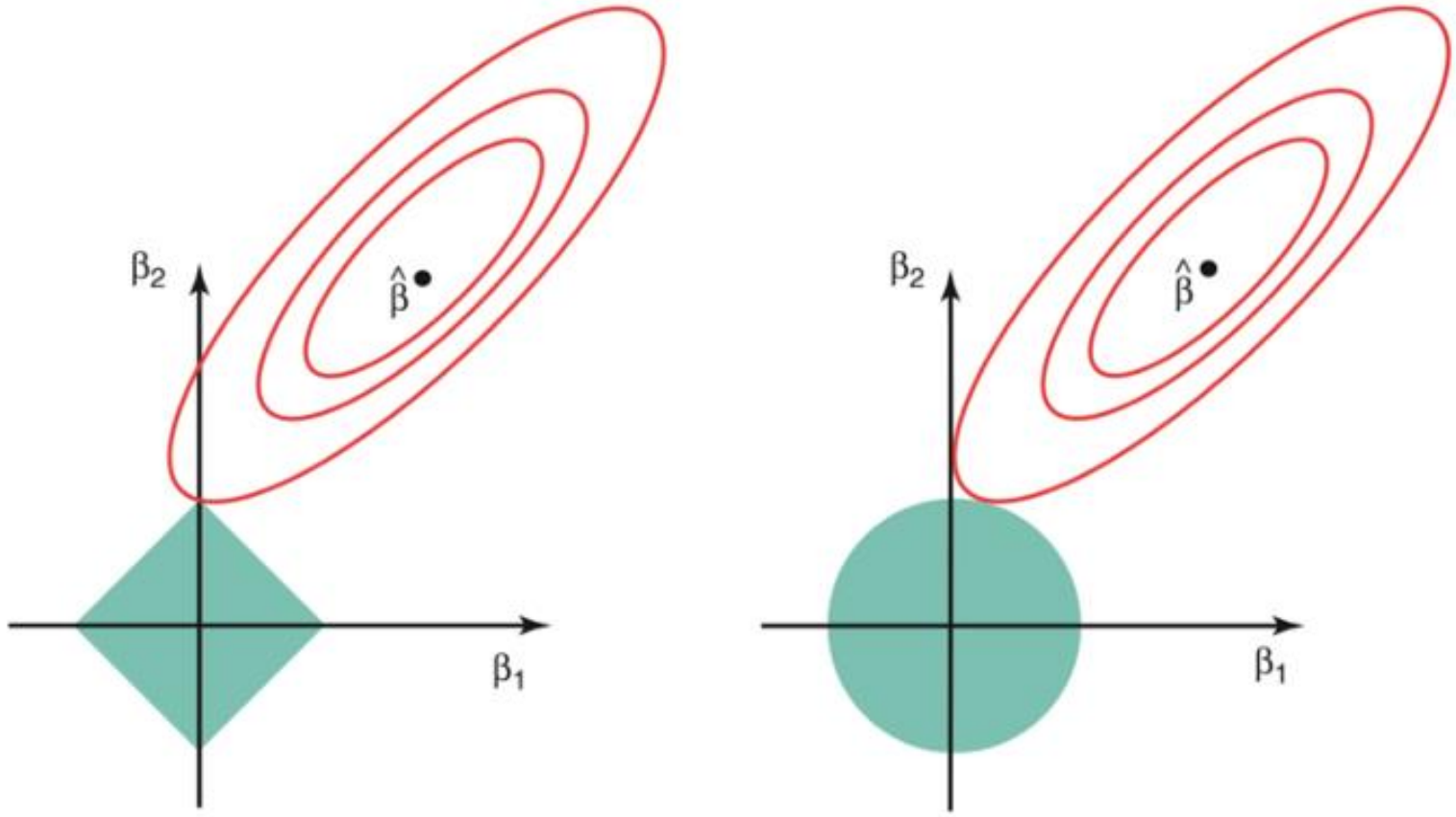
Allianz

Shrinkage Methods

- So, Ridge and Lasso are actually classes of estimators, since they depend on λ
- How to choose the right λ? Cross validation!

Allianz

Shrinkage Methods – geometrical interpretation



Allianz

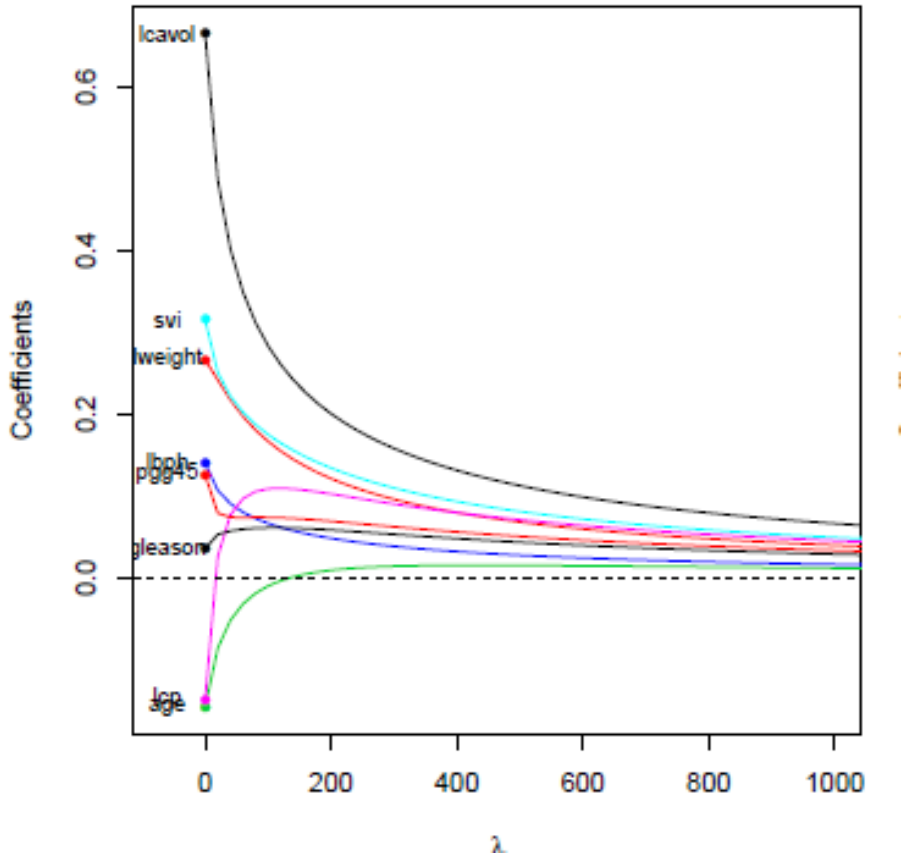
Model selection

- Ridge estimator will almost surely not set any estimated coefficients to zero because of its L2 geometry
- On the other hand, that is exactly what happens with Lasso estimates, because of the L1 norm.
- The larger the λ the more coefficients are set to 0.
- So, Lasso performs **model selection and estimation** at the same time

Example – Prostate data

The more you increase λ, the smaller the estimated coefficients are

Ridge estimated coefficients:



Example – Prostate data

The more you increase λ, the smaller the estimated coefficients are

Lasso estimated coefficients: here they are set to 0 for large λ