Benefits:

-Judging performance (player/team) $value

-Ranking

-Which team better for opposite team.

Evaluate the validity of the data:

* Void
* Accurate
* Complete
* Contains the variables

Feature engineering

1)-take eX, log x, ln, x2, X3, interactions

2)-normalize the data

3) randomize data

4) split ratio

“All models are wrong”

Model building –

Underfitting – decreasing generalization, and train error // llow variance /high bais

Overfitting – decreasing training error, // high varaiance, low bais

Bais – difference between expected value and correct value

Variance -

##regression

An example of a closed form solution in linear regression would be the least square equation

*β*^=(*XTX*)−1*XTy*

Can compute inverse if more observations than variables, N, >k

Order of computing - Ok3

summary(lm(demand~price)) vs glm??

## Elasticnet

## Ridge – reduce to non-zero? (lambda\*sq value of the cooeff)

## LASSO/glmnet – reduce some variables to zero/also feature selection (lambda\*absolute value of coefficient)

(Et\*E = e2+e22+23

Vary lambda – to get the smallest root mean square error

Data sources:

Basketballreference.com

* Game, play, and team data
* Distance traveld to play
* City coordinates(wiki)
* City GDp (Wikipedia +US Govt’)
* Ratings – sonny momore, jeff Sagarin

R:

rm(list=ls())

par(mfrow=c(2,2))