Regression

Carlos Soares

(partly using materials from Moreira, Carvalho & Horvath)

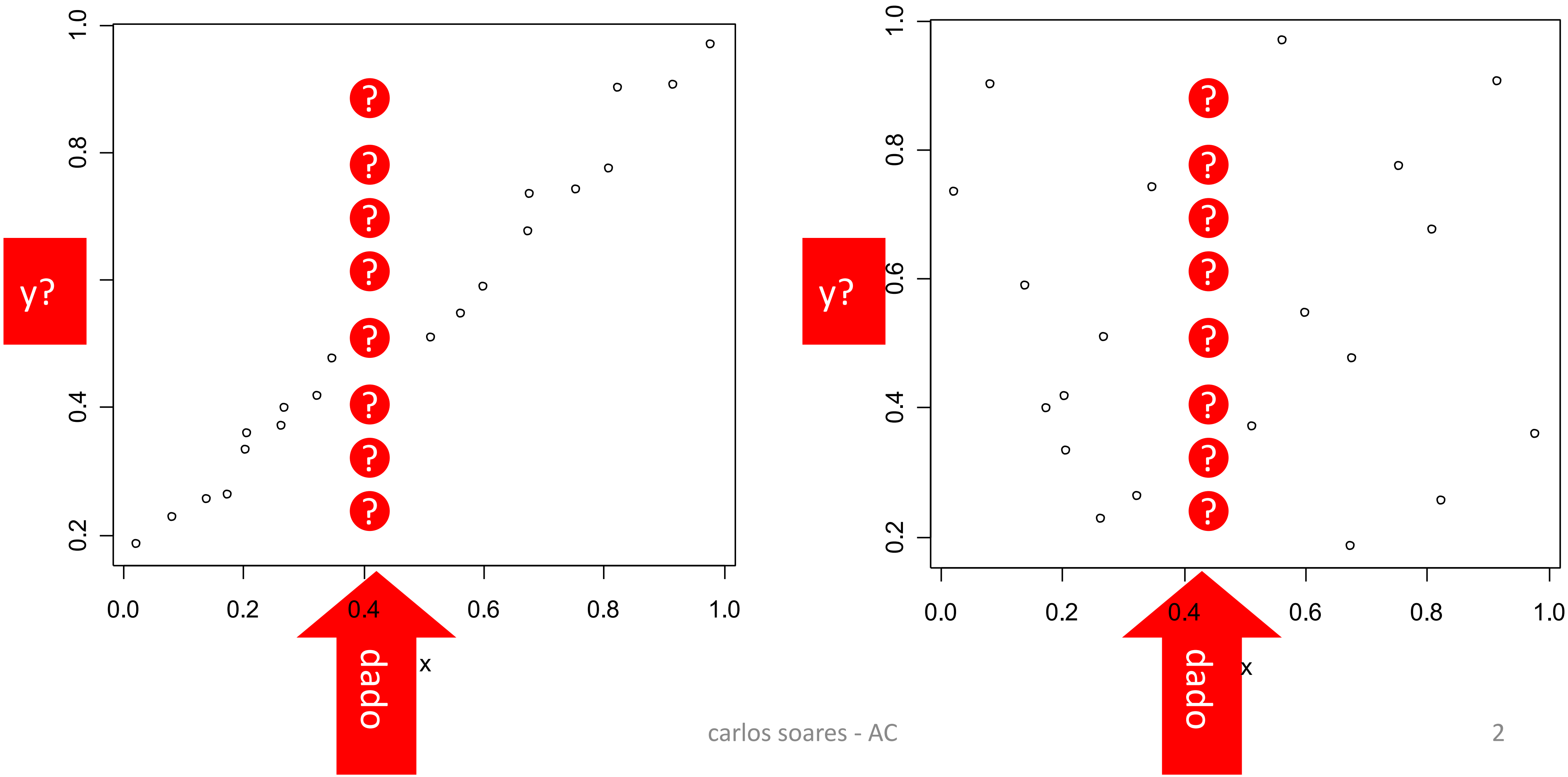
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

http://www.flickr.com/photos/upv
cfrontdoor/4642103837/

regression

x = family income

y = total purchases



plan & goals

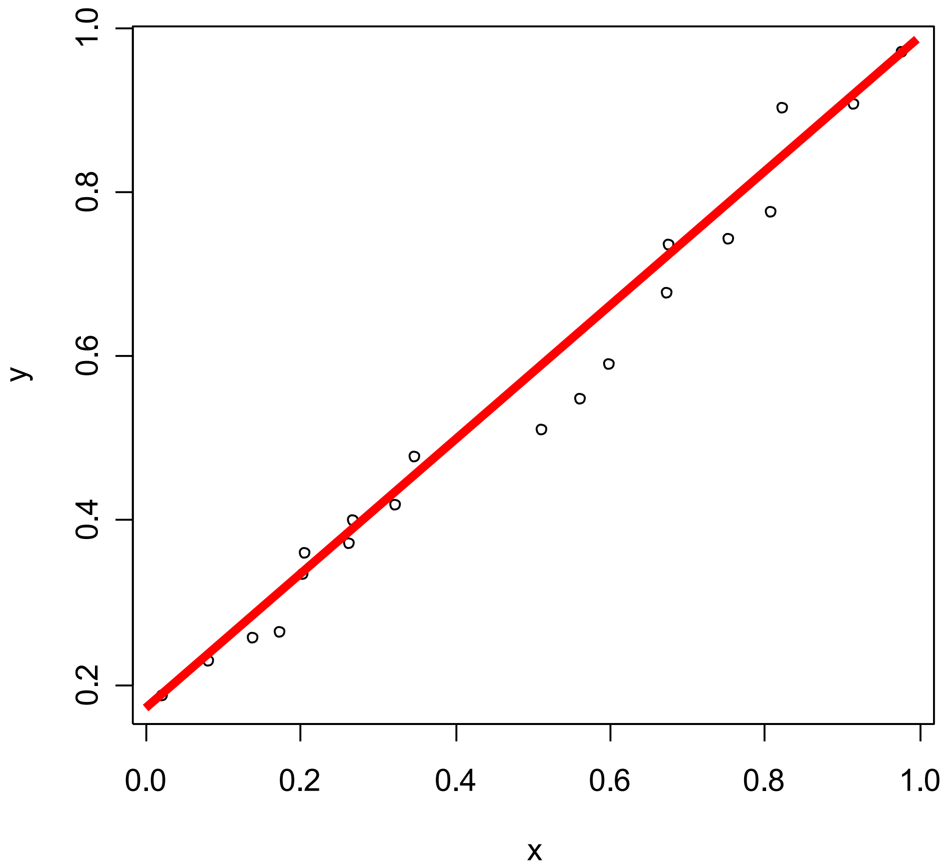
- linear regression
  - interpretation
  - algorithm
- evaluation of regression models
- other algorithms
- regression concepts
  - interpretation of the linear model
  - evaluation measures
- common approaches to dapting learning algorithms for regression

linear regression

- simple case: 2 variables
  x and *y*

- liner equation

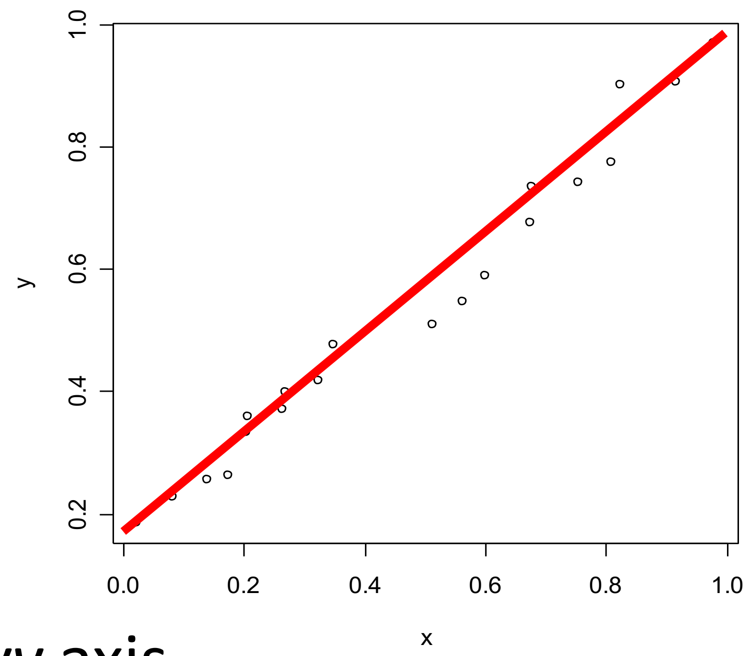
$$\begin{aligned} y &= f(x) \\ &= b_0 + b_1 x \end{aligned}$$



interpretation of coefficients

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

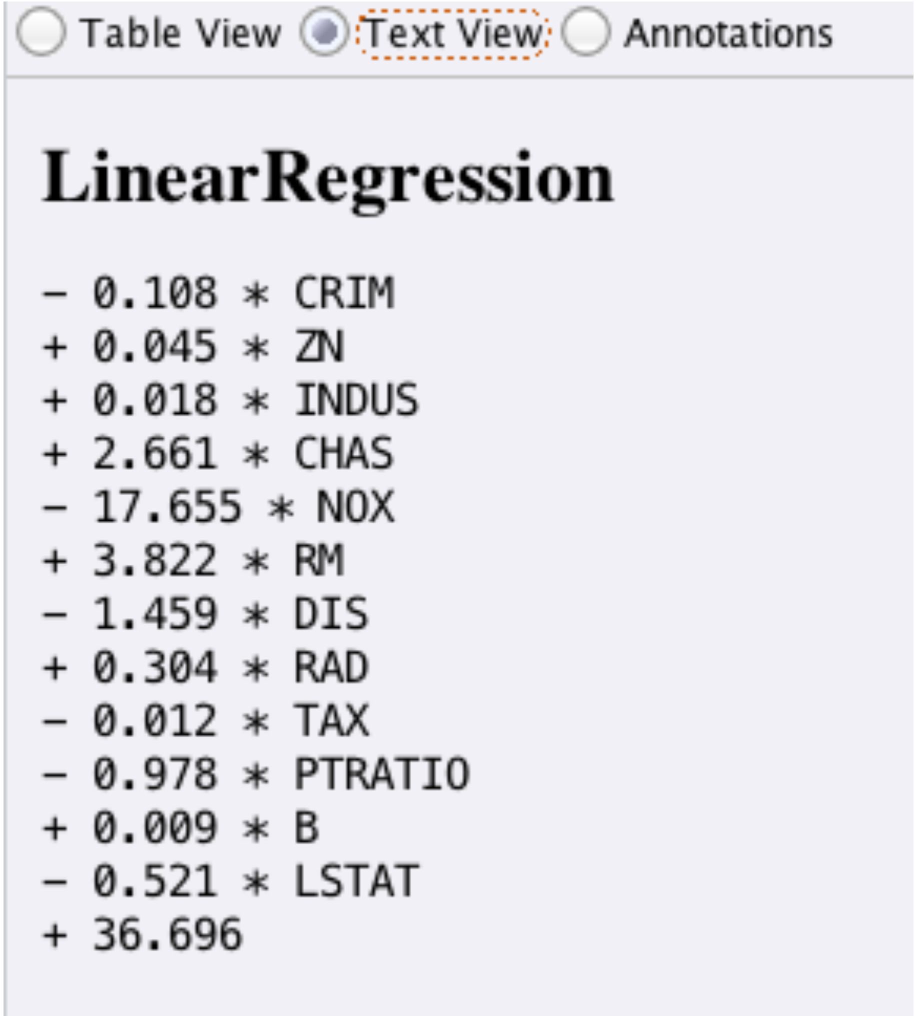
$$y = b_0 + b_1 x$$



- b_0: intersection of the line with the *yy* axis
  - often hard to interpret
- b_1: slope of the line
  - variation in the value of *y* given a 1 unit increase of the value of *x*

exercise II: analyze linear regression model

- assumes that variables are not correlated
  - influence of each variable is explained separately
  - coefficients are not influenced by changing the set of explanatory variables
    - i.e. attributes
- variation depends on the degree of correlation
  - signal may change!
- ... but empirical results show robustness



Simple linear regression: estimating parameters

$$y = b_0 + b_1 x$$

$$\widehat{b_1} = \frac{S_{XY}}{S_{XX}}$$

where $\widehat{b_1}$ is an estimate of b

$$S_{XY} = \sum_{i=1}^{n} [(X_i - \overline{X}).(Y_i - \overline{Y})]$$

$$S_{XX} = \sum_{i=1}^{n} (X_i - \overline{X})^2$$

- $\widehat{b_1}$ should be statistically significantly different from zero
  - if not, there is no meaningful dependency between *Y* and X
  - this should be tested

$$\widehat{b_0} = \overline{Y} - \hat{\beta}.\overline{X}$$

where $\widehat{b_0}$ is an estimate of b_0

- $\widehat{b_0}$ may or may not be statistically significantly different from zero
  - If not there is no evidence that *Y≠0* when *X=0.*
  - … which could make sense
    - e.g. value of a customer with 0 income
  - … or not…
    - e.g. minimum sales of a product without shelf space

Simple linear regression: assumptions

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

- Linear relationship between *x* and *y*
  - also additive
- Errors
    - i.e. unexplained variation in *y*
  - ... are independently and identically distributed
  - ... homoscedasticity
    - constant variance
  - ... normally distributed

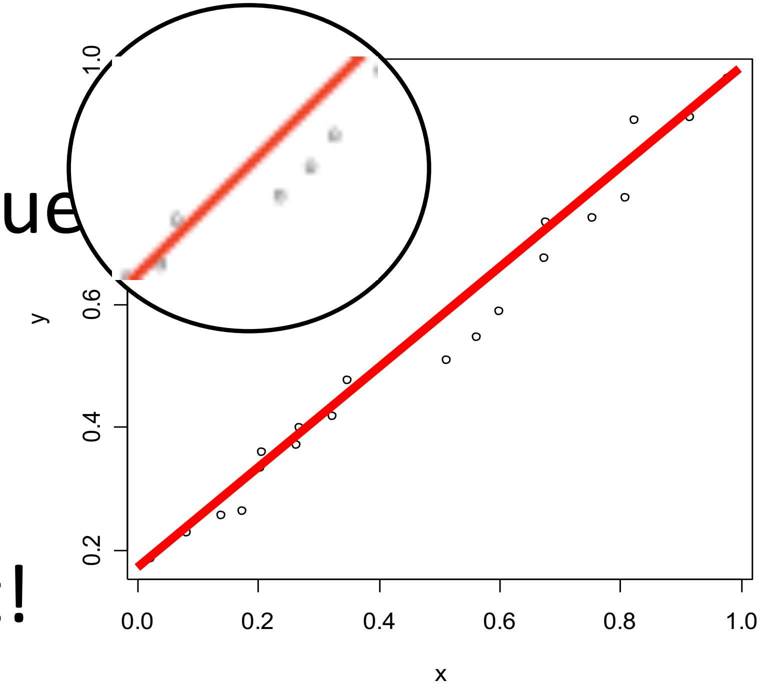
gps

- linear regression
- evaluation of regression models
  - measures
  - methodology
  - bias-variance trade-off
- other algorithms

prediction and evaluation

- given the value of *x*
- ... the model estimates the value of *y*

$$\hat{y} = b_0 + b_1 x$$



- but the estimate is not perfect!

- erro:
  - *y* : true value
  - $\hat{y}$: value estimated by the model

$$\hat{y} - y$$

analysis of evaluation measures

- mean error
  - DO NOT USE!

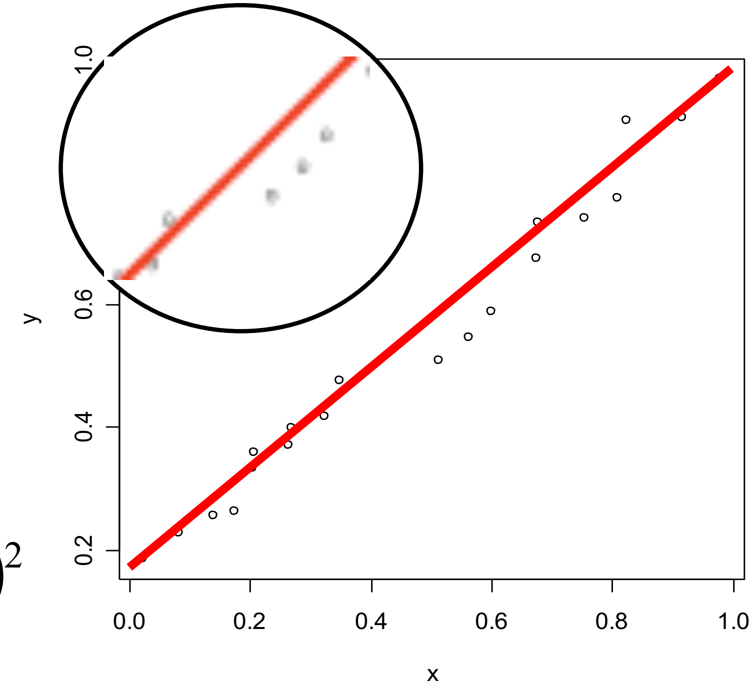
$$\frac{1}{m}\sum_i \hat{y}_i - y_i$$

- mean absolute error
  - estimates "typical" error

$$\frac{1}{m}\sum_i \left|\hat{y}_i - y_i\right|$$

- mean squared error
  - assigns more weight to larger errors
  - ... may be dominated by a few cases

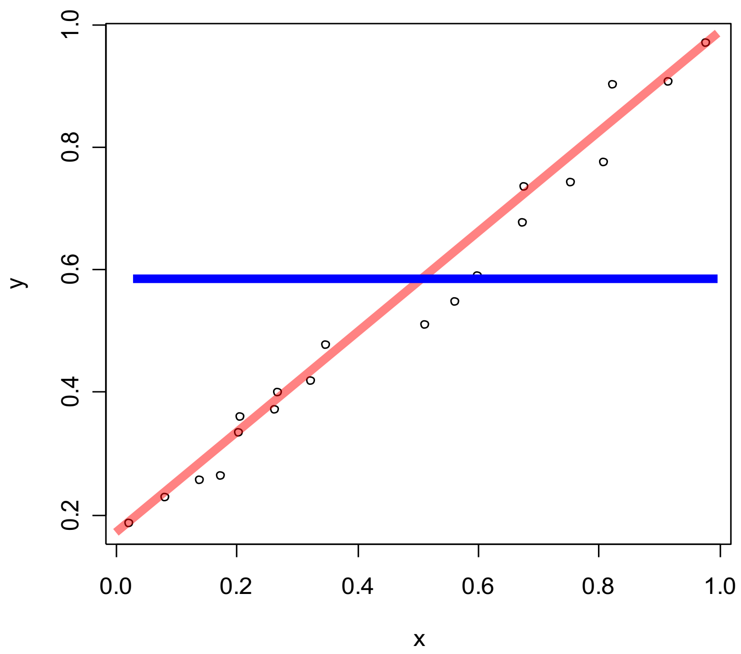
$$\frac{1}{m}\sum_i \left(\hat{y}_i - y_i\right)^2$$



- values depend on the scale of the target variable
  - is the error good or bad?
    - business perspective?
    - does the relationship between x/y represented really exist?

baseline: trivial model

- if we know nothing about the cases
- what is the best prediction we can make?
  - random vs mean



- trivial model $\hat{y}_i = \bar{y}$

- regression is only useful if its error is lower than the one obtained with the triveal prediction
  - eg. mean squared error

$$\frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2}$$

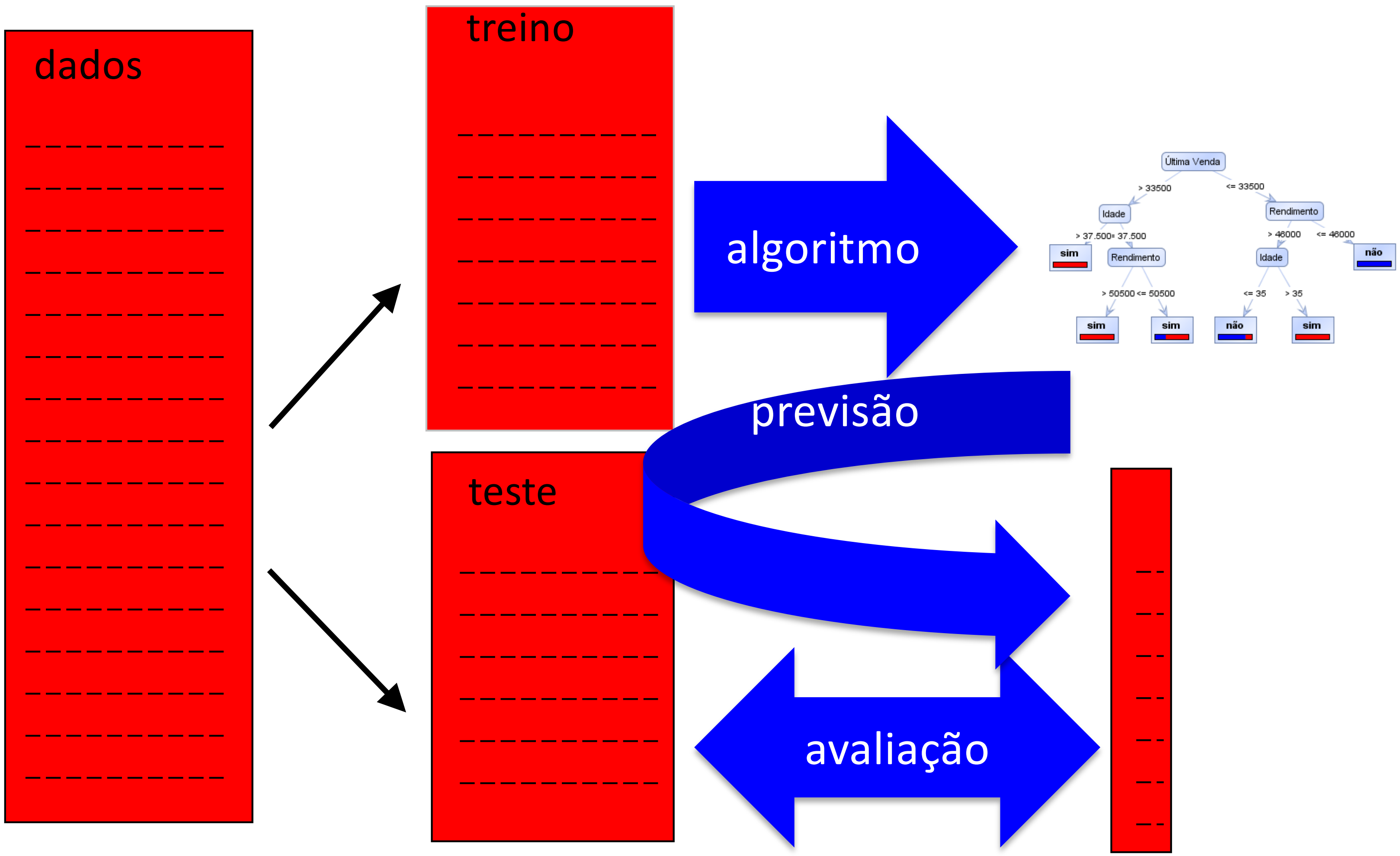
0 if regression model is perfect

]0,1[if it is useful

1 if it is equivalent to the trivial model

>1 if it is worse than the trivial model

evaluation methodology: do not forget!

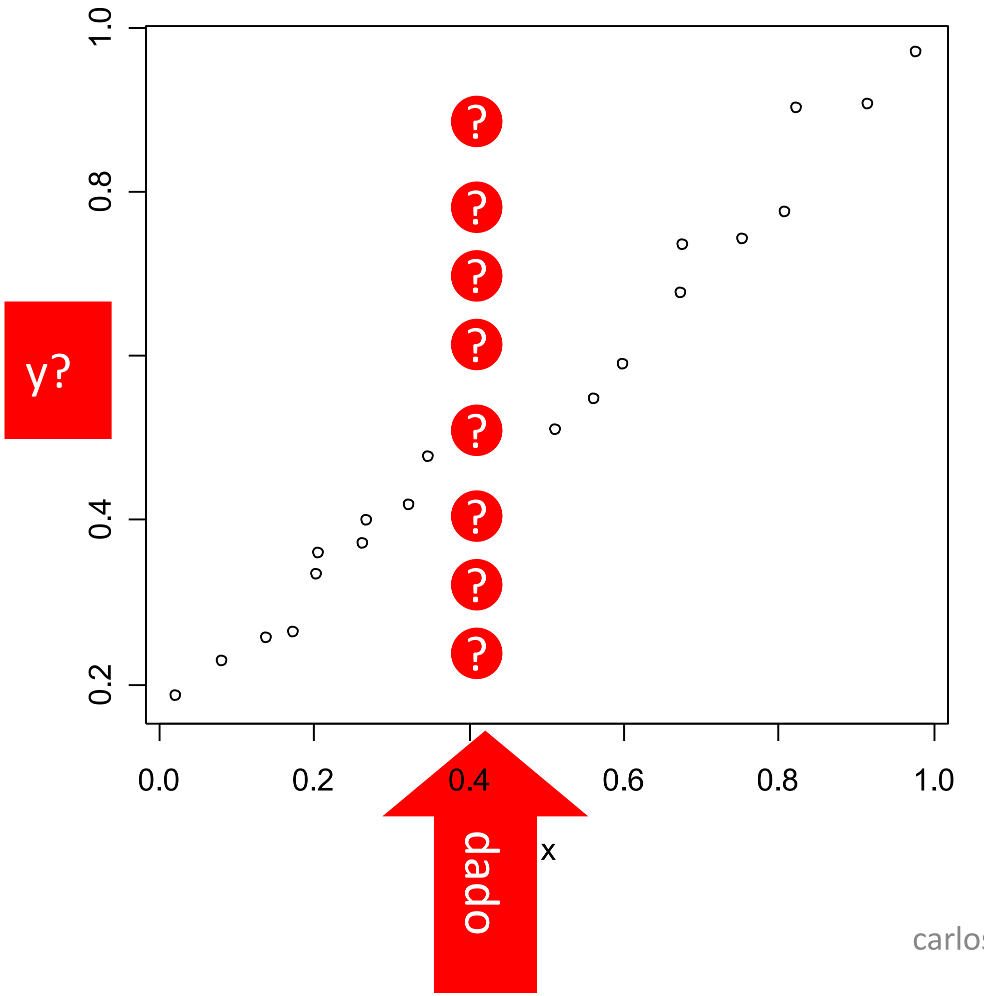


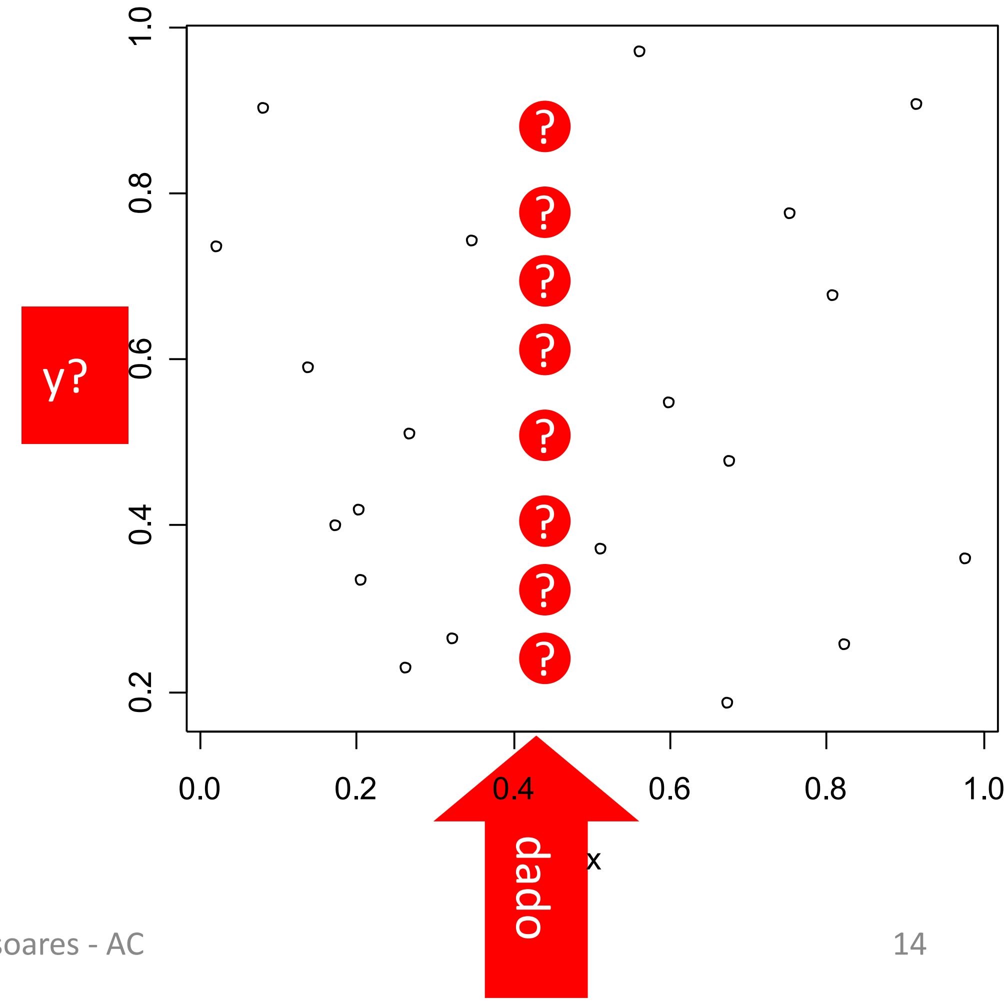
ex. MSE

remember?

x = family income

y = total purchases



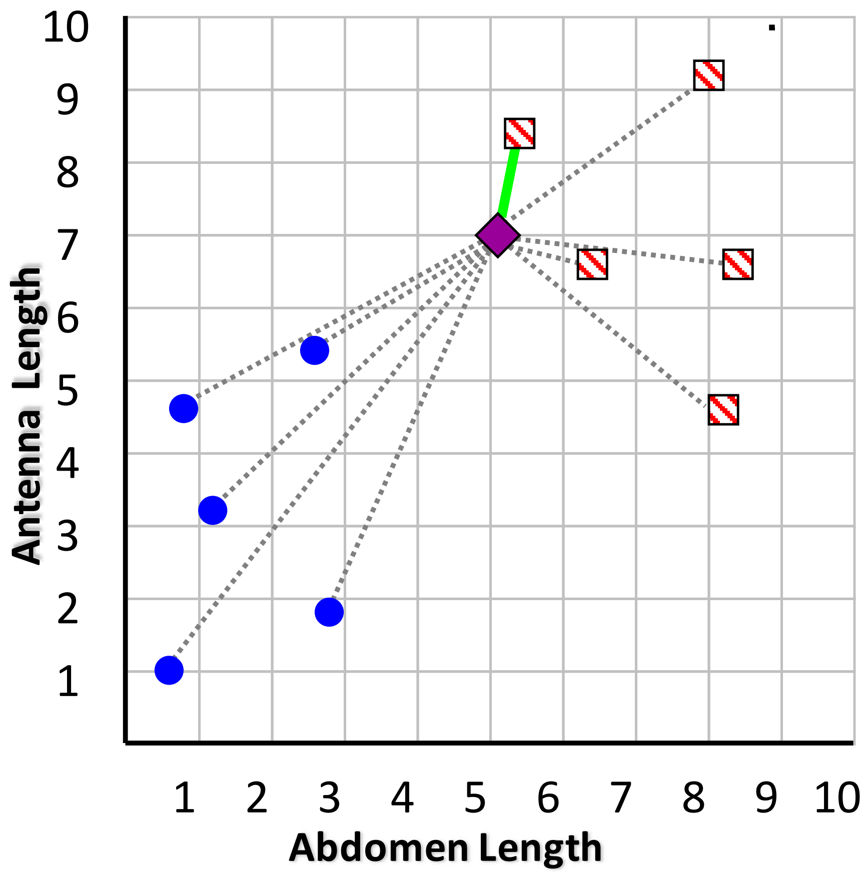


gps

- linear regression
- evaluation of regression models
- other algorithms
  - kNN
  - trees
  - neural networks
  - support vector machines
  - ... bias & variance

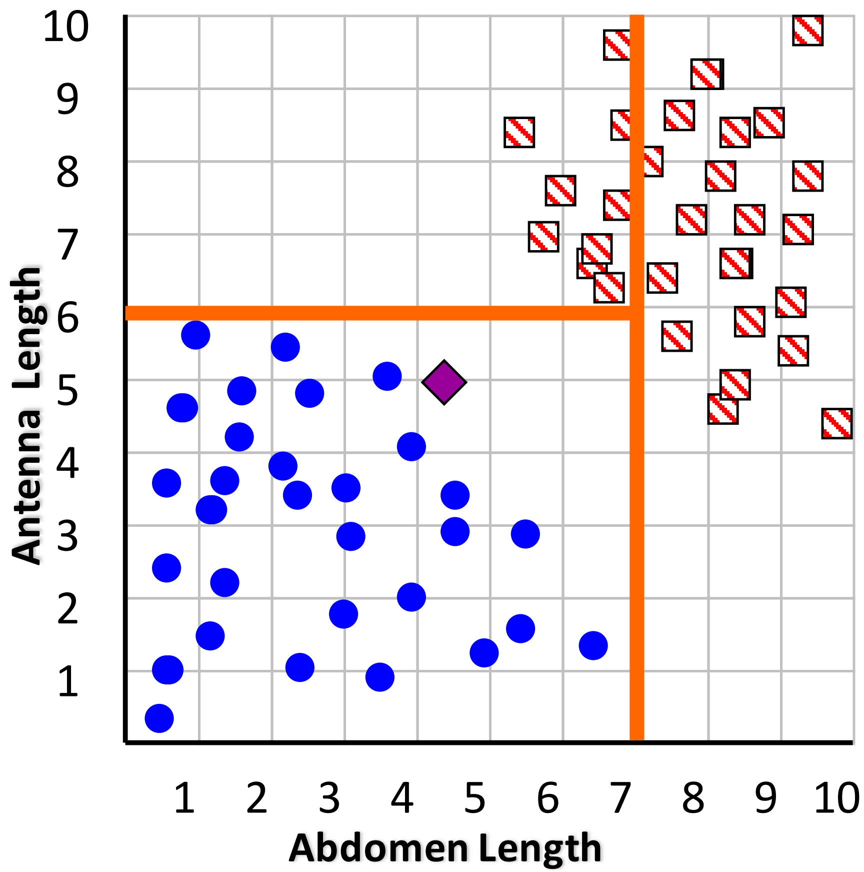
Nearest Neighbor Algorithm for Regression

- find kNN
  - just like for classification
- predict the average of their target values
  - instead of majority voting



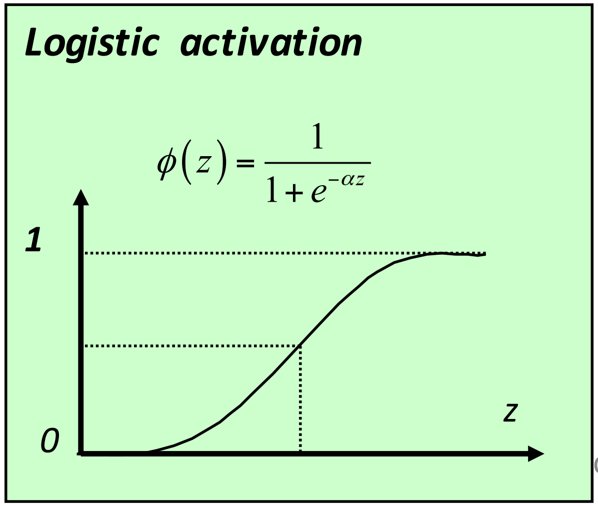
Decision Trees for Regression

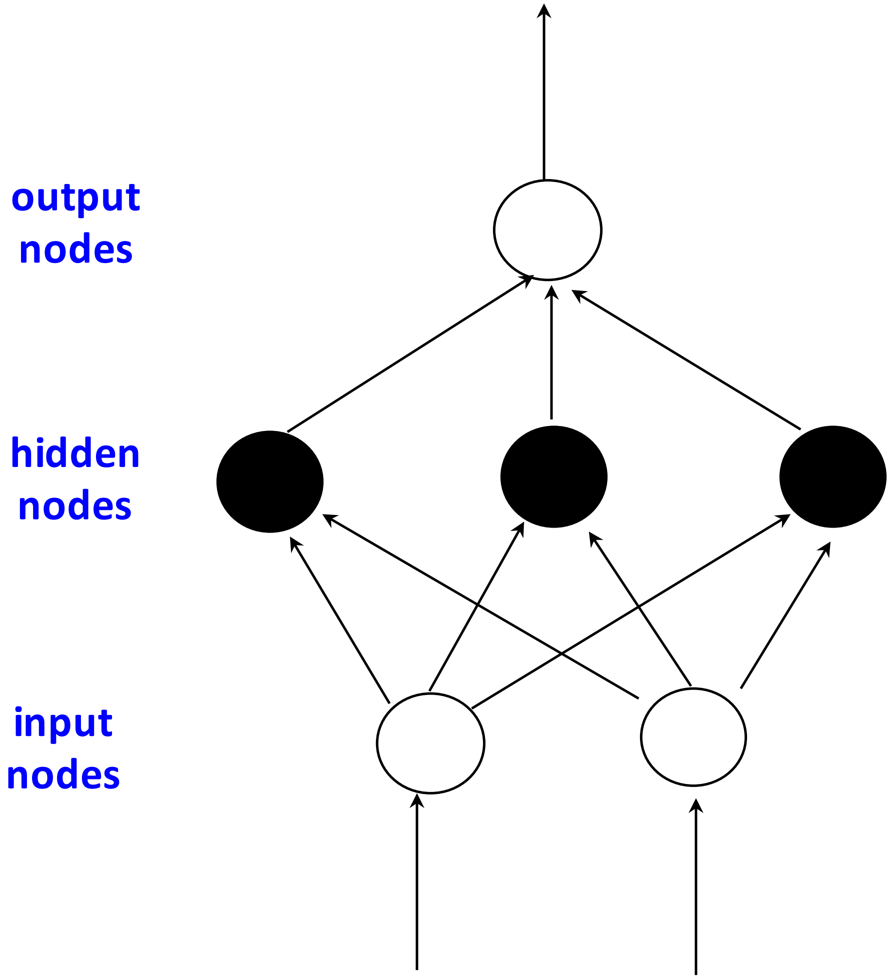
- train
  - splitting criterion based on the sum of the variances
    - instead of gini or entropy
- prediction
  - average of targets in the leaf
    - instead of majority voting
- variants
  - model trees
    - msing MLR or K-NN in the leaves instead of the average
  - MARS
    - multivariate adaptive regression splines



Neural Nets for Regression

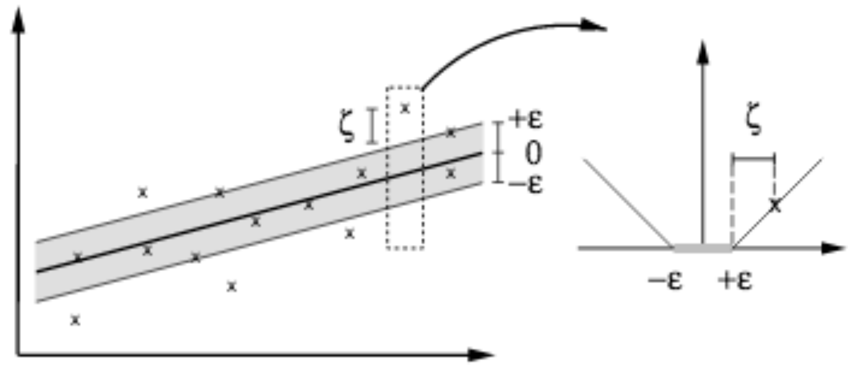
- single output node
  - predicted y = score
- continuous activation function
  - e.g. sigmoid
    - also used for classification





SVM for Regression

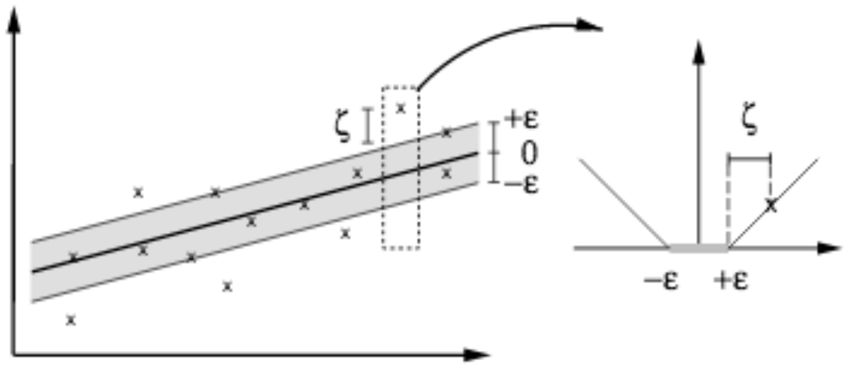
- margin
  - minimize the tube "around" the data
    - Instead of maximizing the distance to closest examples from each class



source: http://alex.smola.org/papers/2003/SmoSch03b.pdf

SVM for Regression

- margin
  – minimize the tube "around" the data
    - Instead of maximizing the distance to closest examples from each class



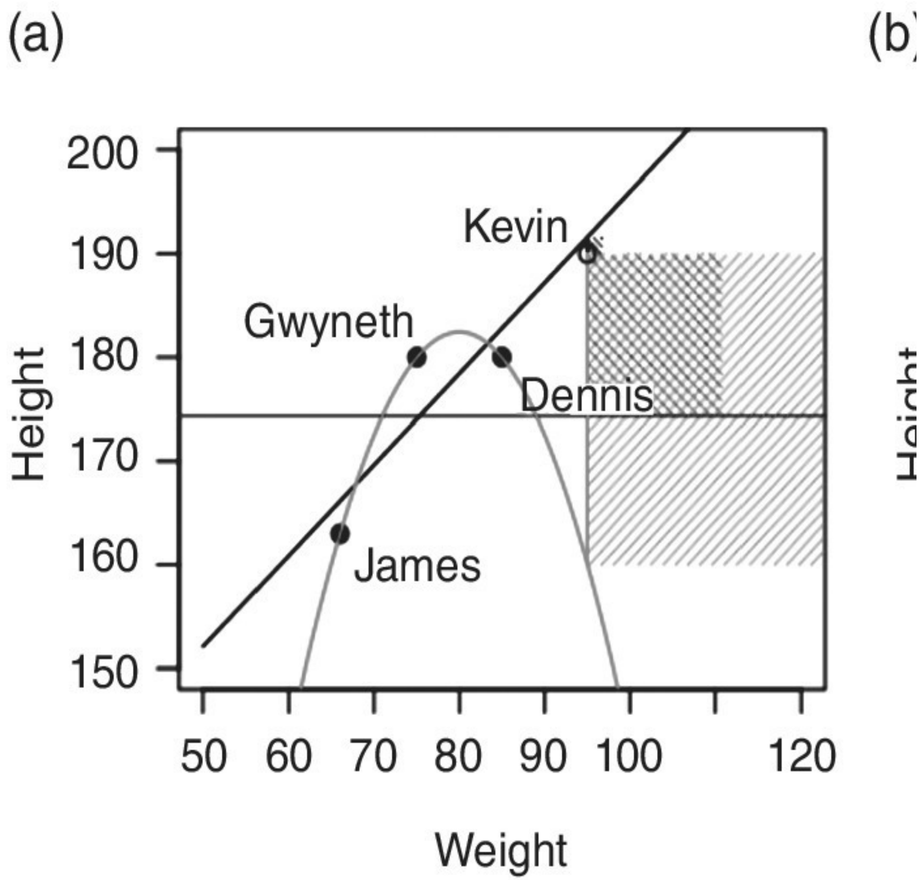
source: http://alex.smola.org/papers/2003/SmoSch03b.pdf

bias

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

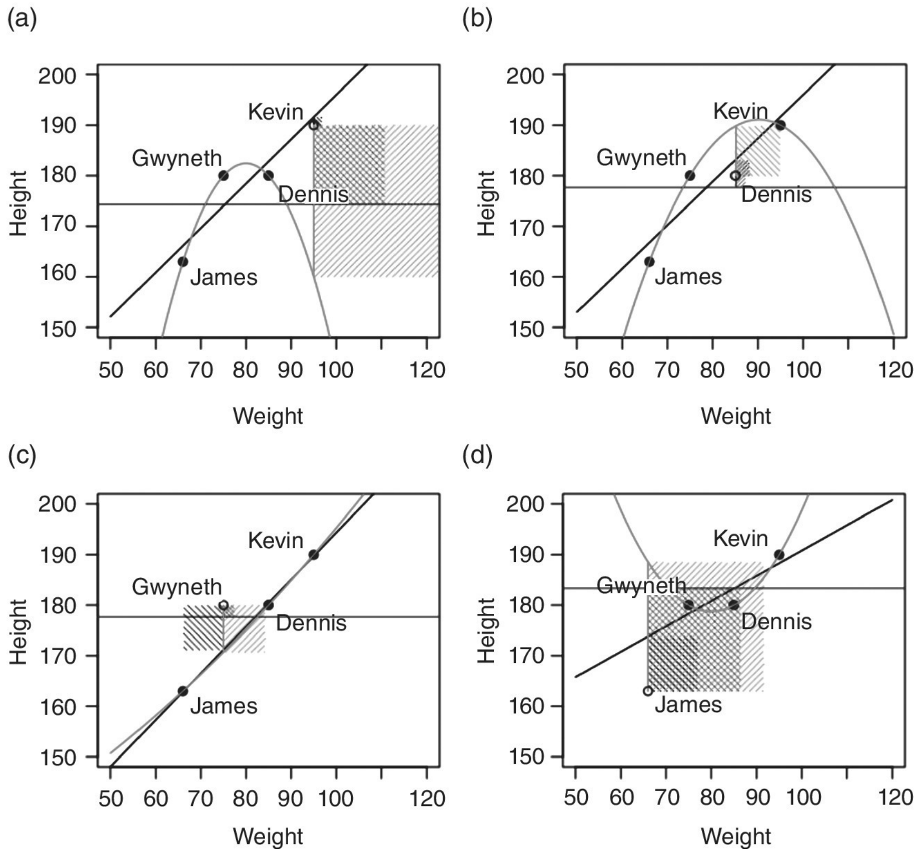
- type of model an algorithm is able to learn given a set of training data

- related to hypothesis language
  - e.g. linear vs quadratic



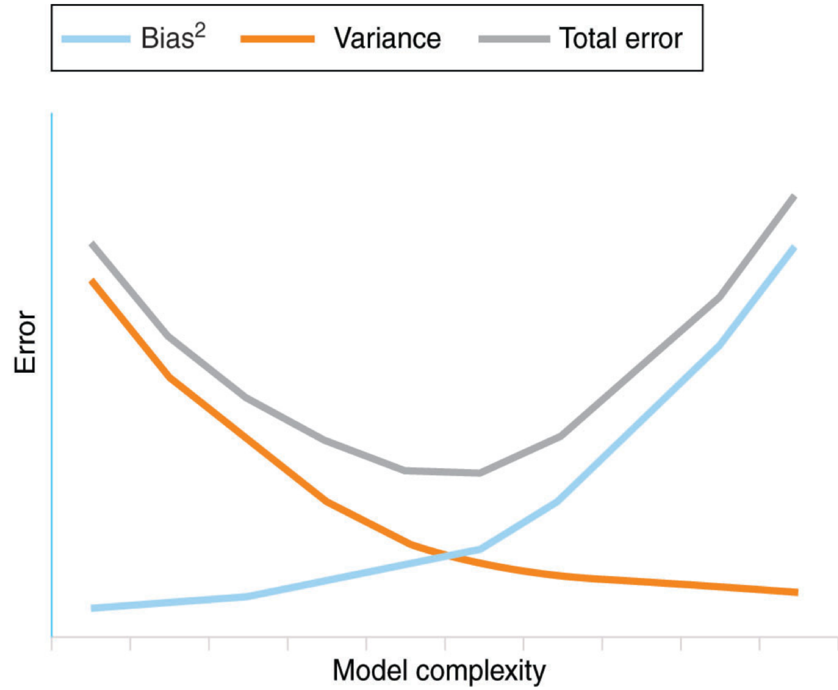
… and variance

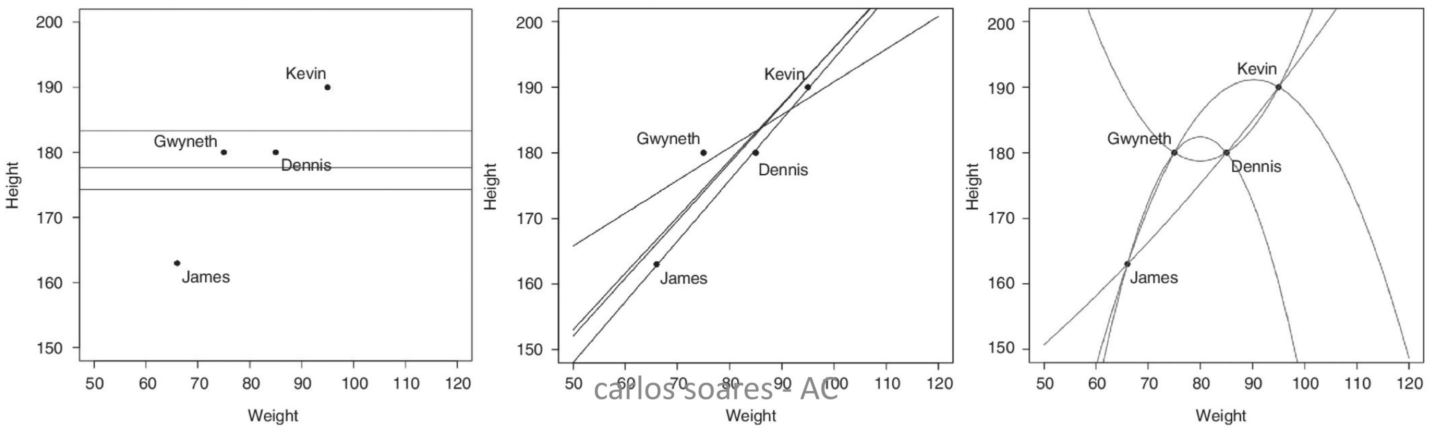
- variation in model an algorithm is able to learn, given different training data
  - ie. small changes



bias-variance trade-off

- Low bias implies high variance and vice-versa
- We would like to find a model with a good trade-off
  - Not too complex but with good predictive power

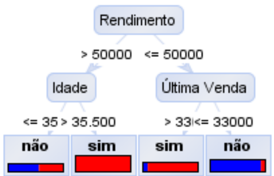




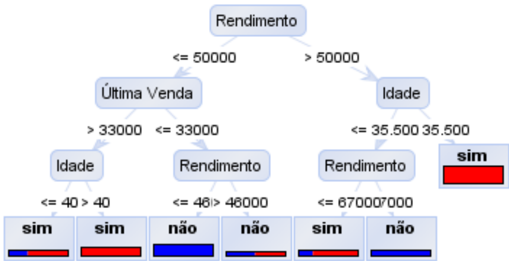
remember overfitting?

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

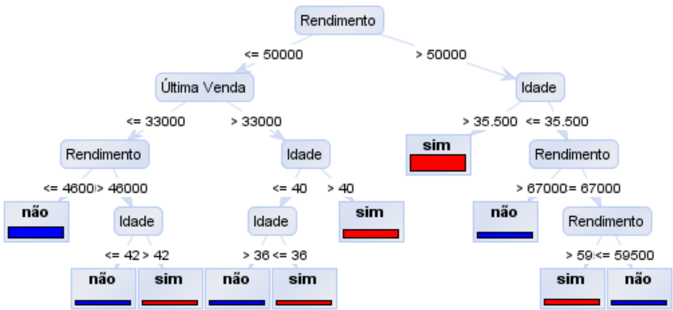
- trees obtained with different values of “minimum leaf size”
  - 4, 2 and 1



error (train)=18,18%



error (train)=9,09%



error (train)=0,00%