Regression

Carlos Soares

(partly using materials from Moreira, Carvalho & Horvath)

U. PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

http://www.flickr.com/photos/upvcfrontdoor/4642103837/

reference materials

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

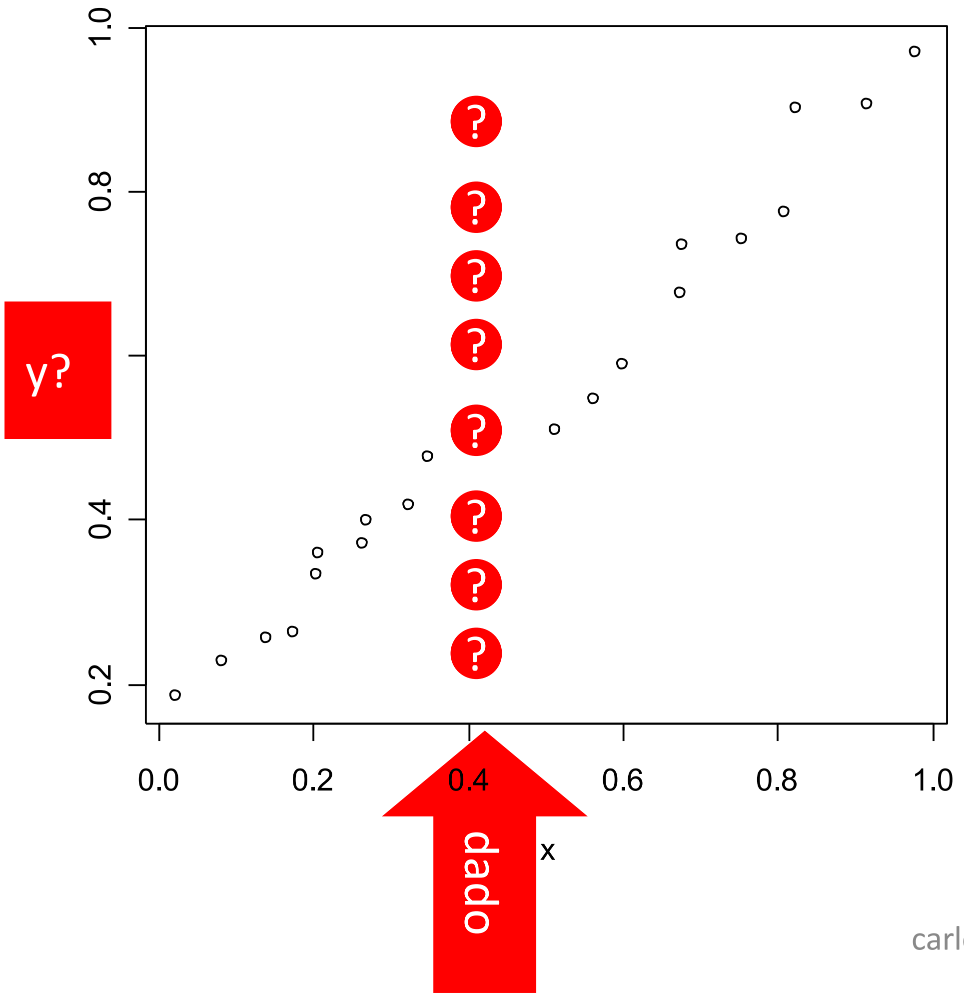
- JMM et al. ch. 8+9+10+12 (parts)

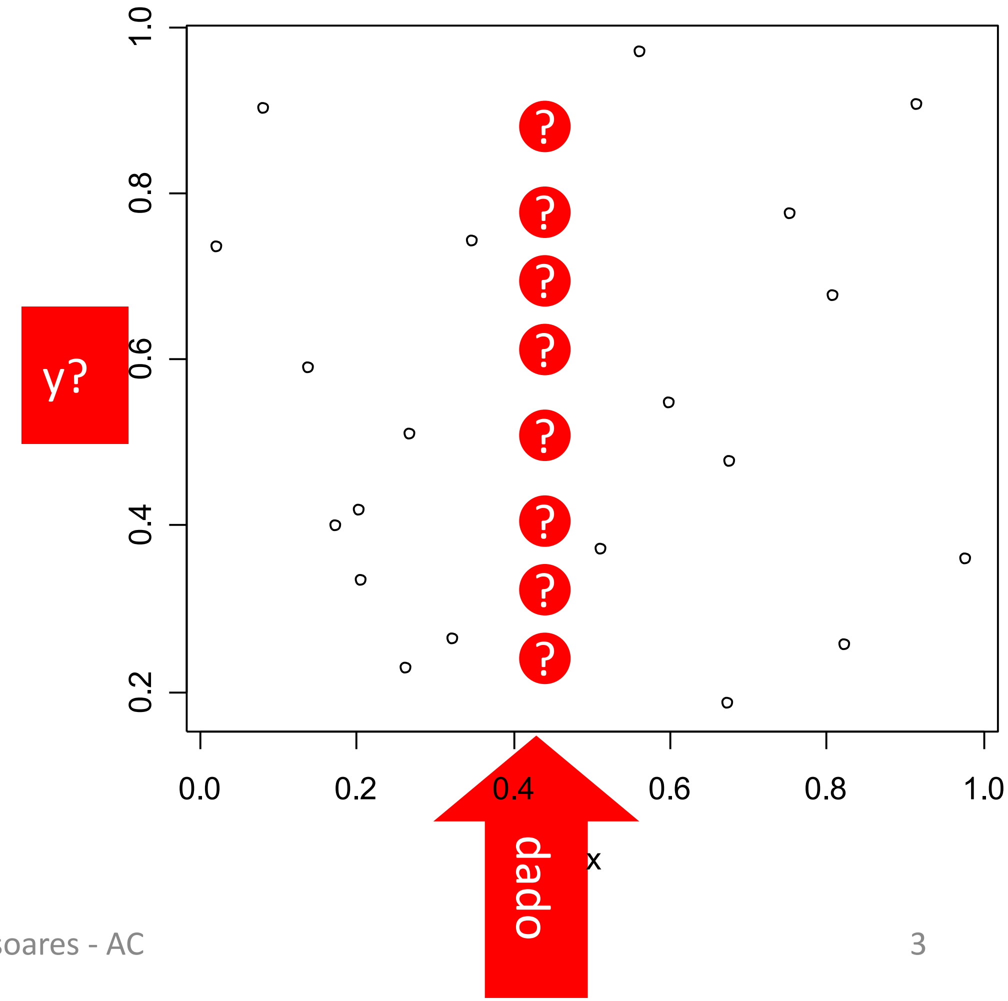
regression

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

x = family income

y = total purchases





plan & goals

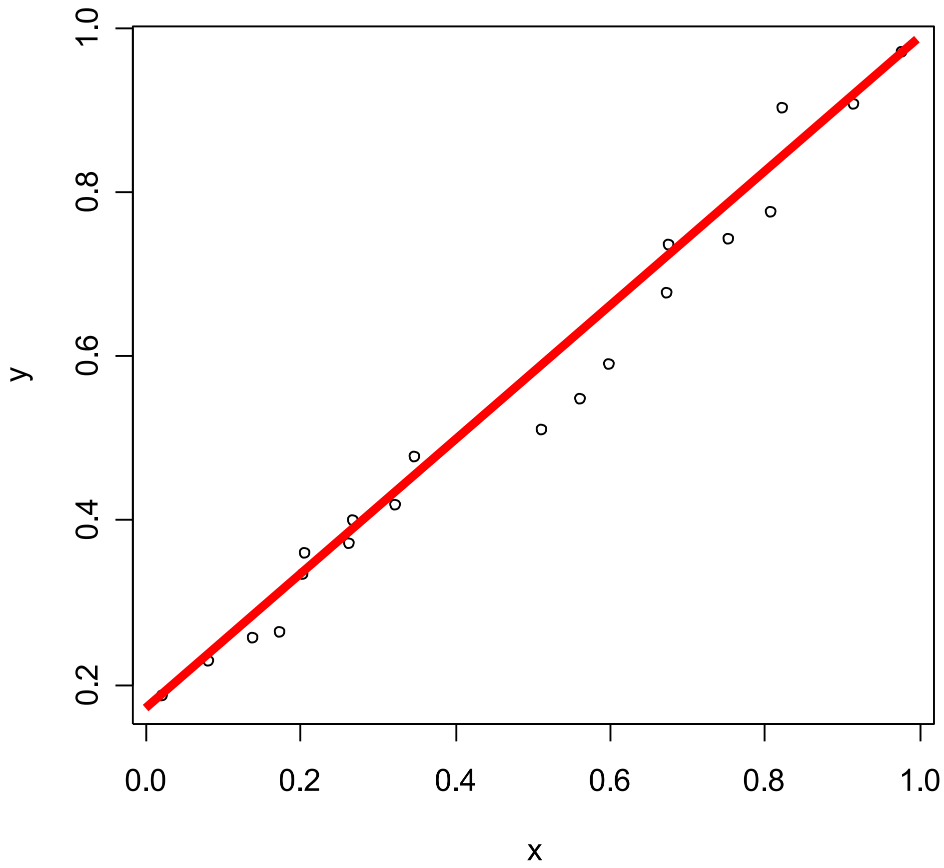
- linear regression
  - interpretation
  - algorithm
- evaluation of regression models
- other algorithms
- bias & variance decomposition of error
- regression concepts
  - interpretation of the linear model
  - evaluation measures
- common approaches to dapting learning algorithms for regression

linear regression

- simple case: 2 variables
  x and *y*

- liner equation

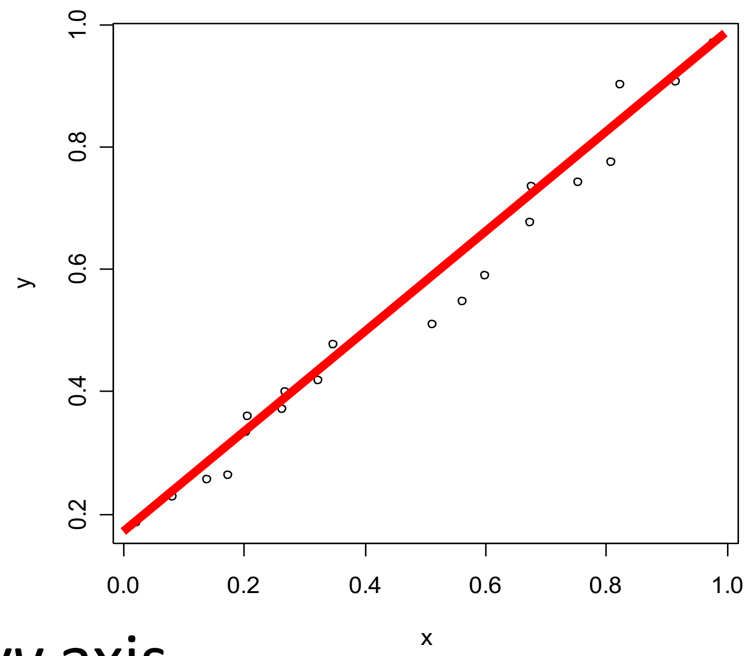
$$
\begin{aligned}
y &= f(x) \\
  &= b_0 + b_1 x
\end{aligned}
$$



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*

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

Table View | Text View | Annotations

LinearRegression

```
- 0.108 * CRIM
+ 0.045 * ZN
+ 0.018 * INDUS
+ 2.661 * CHAS
- 17.655 * NOX
+ 3.822 * RM
- 1.459 * DIS
+ 0.304 * RAD
- 0.012 * TAX
- 0.978 * PTRATIO
+ 0.009 * B
- 0.521 * LSTAT
+ 36.696
```

gps

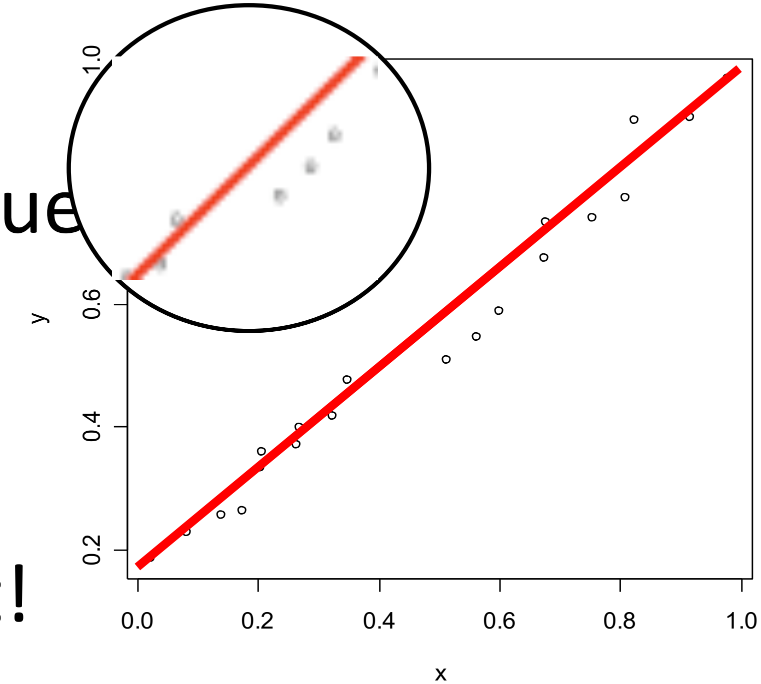
- linear regression
- evaluation of regression models
  - measures
  - methodology
  - bias-variance trade-off
- other algorithms
- bias & variance decomposition of error

prediction and evaluation

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

- 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!

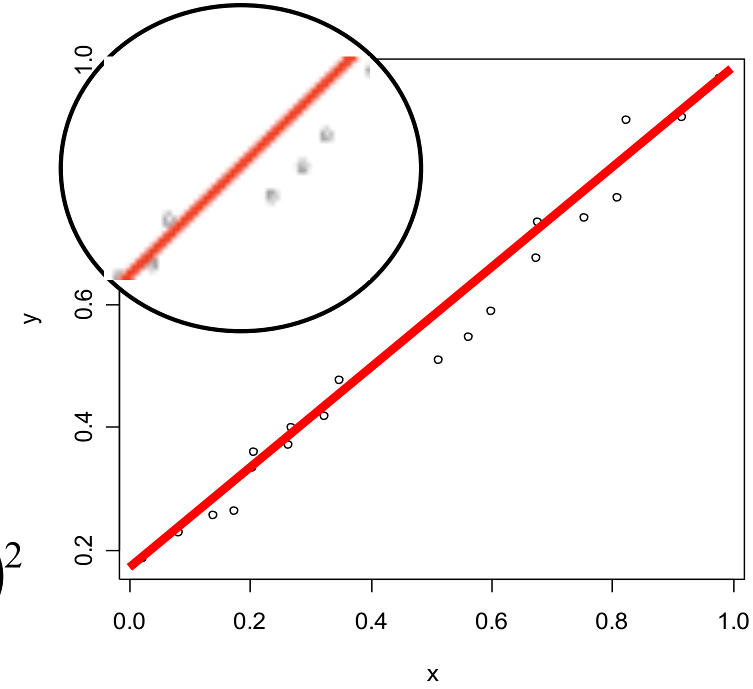
- erro:
  - *y* : true value
  - $\hat{y}$: value estimated by the model

$$\hat{y} - y$$

analysis of evaluation measures

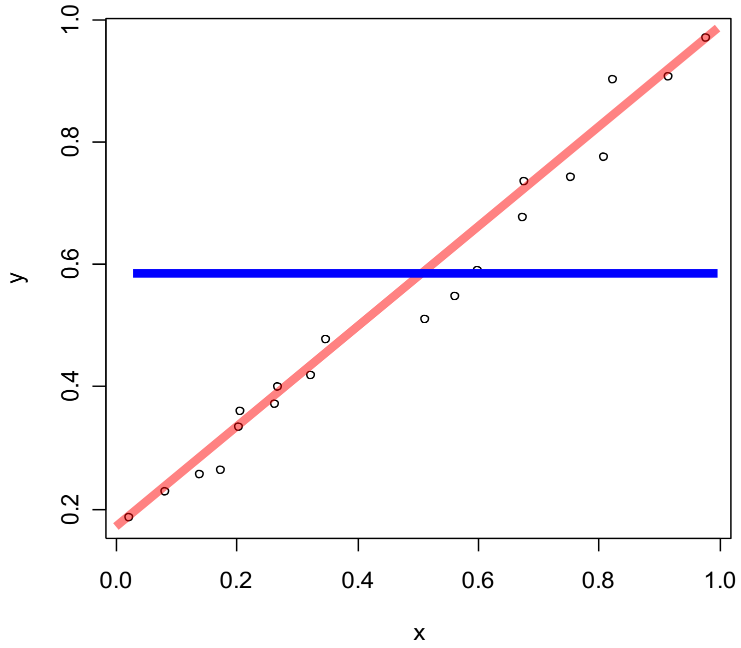
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

- mean error $\frac{1}{m}\sum_i \hat{y}_i - y_i$
  - DO NOT USE!
- mean absolute error $\frac{1}{m}\sum_i \left|\hat{y}_i - y_i\right|$
  - estimates "typical" error
- mean squared error $\frac{1}{m}\sum_i \left(\hat{y}_i - y_i\right)^2$
  - assigns more weight to larger errors
  - ... may be dominated by a few cases
- 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

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

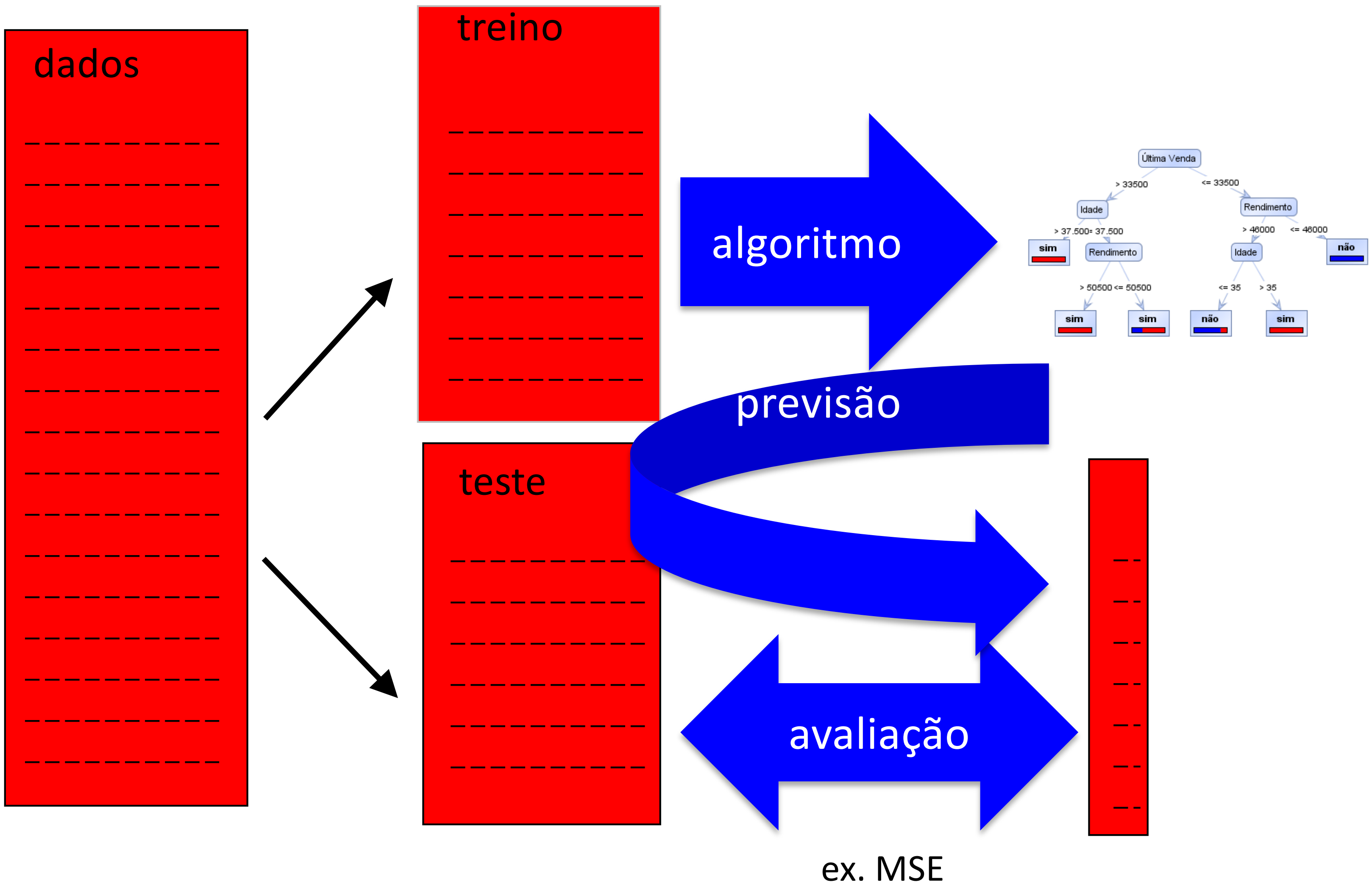
0 if regression model is perfect

]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!



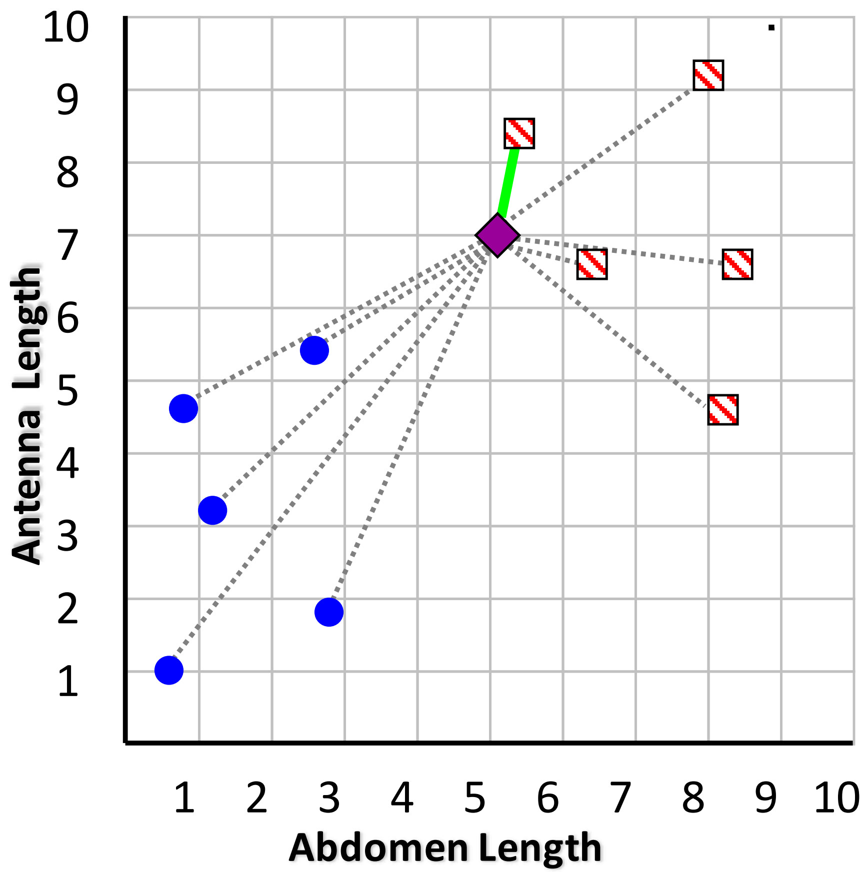
gps

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

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

Nearest Neighbor Algorithm for Regression

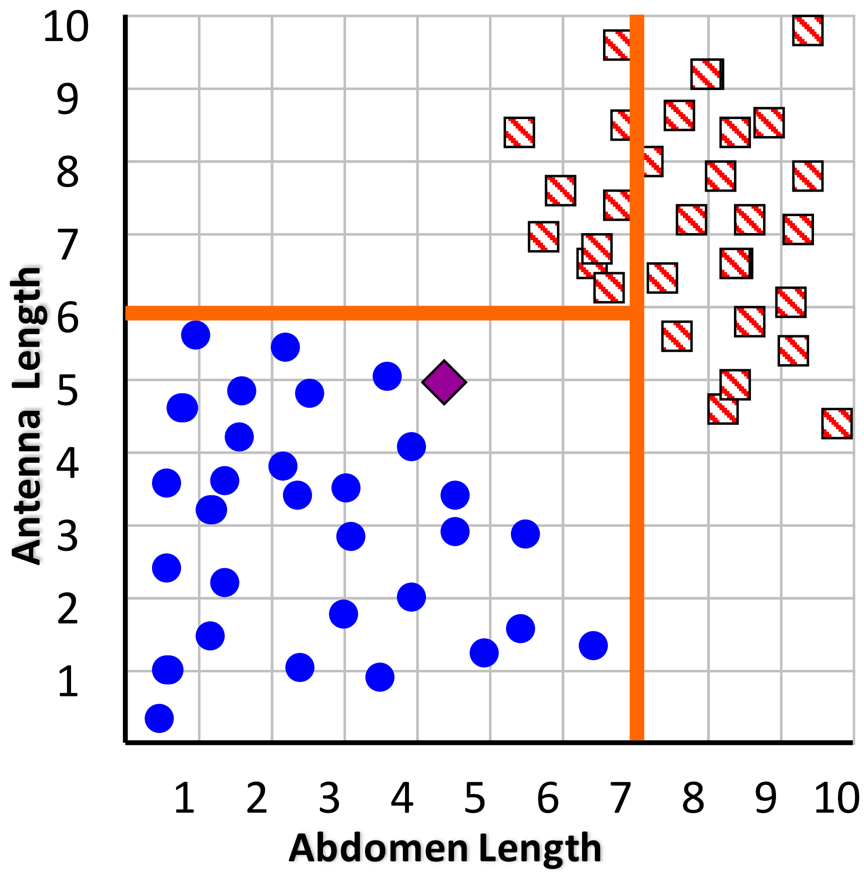
- find kNN
  - just like for classification
- predict the average of their target values
  - instead of majority voting



Decision Trees for Regression

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

- 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
    - using MLR or K-NN in the leaves instead of the average
  - MARS
    - multivariate adaptive regression splines



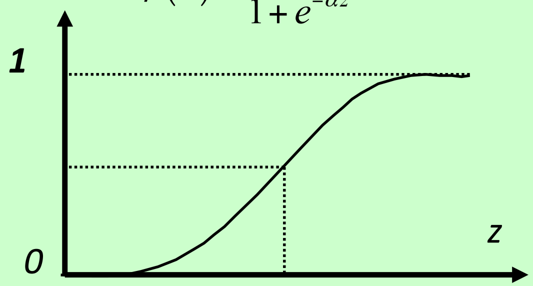
Neural Nets for Regression

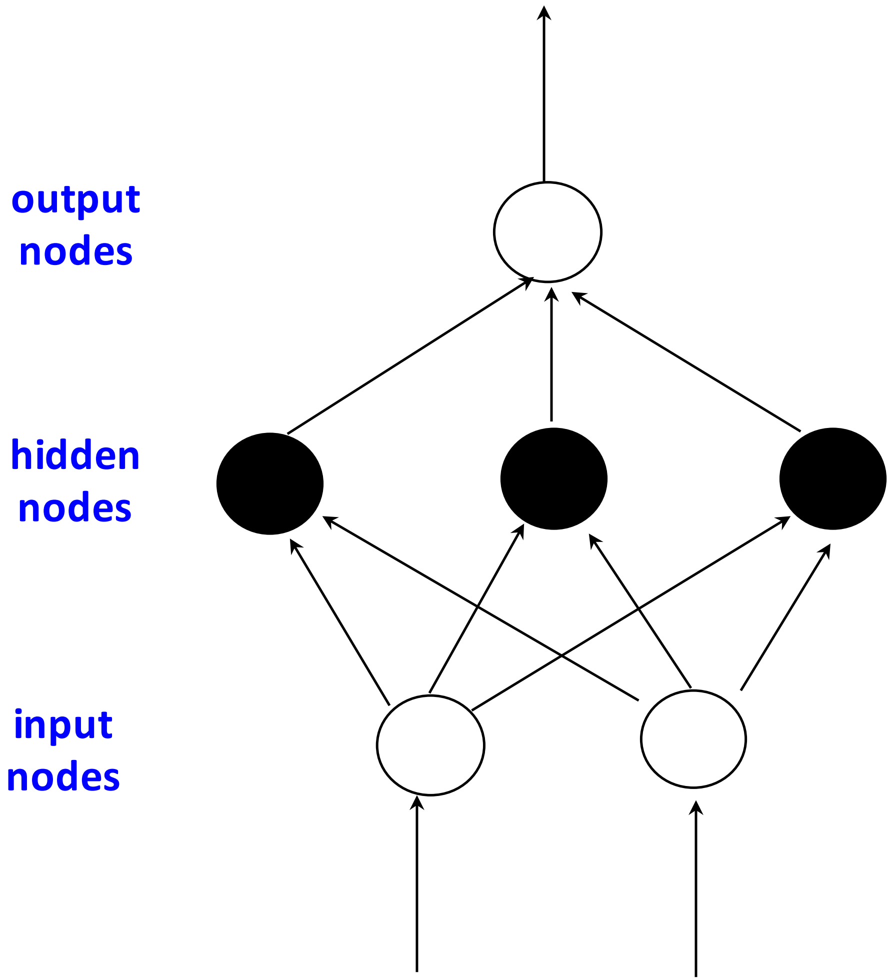
U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

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

Logistic activation

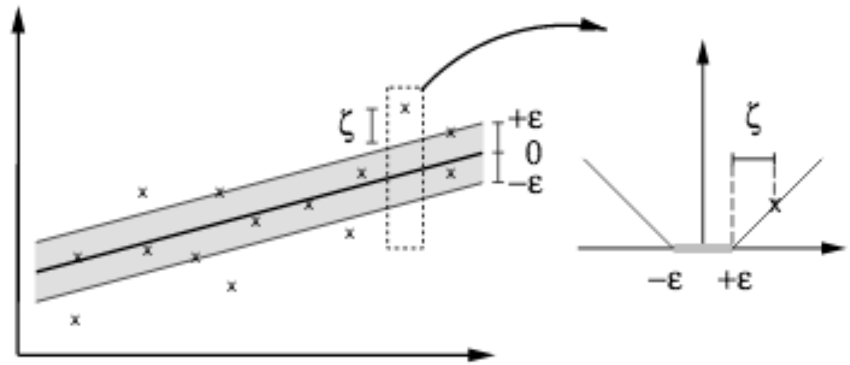
$$\phi(z) = \frac{1}{1+e^{-\alpha z}}$$





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

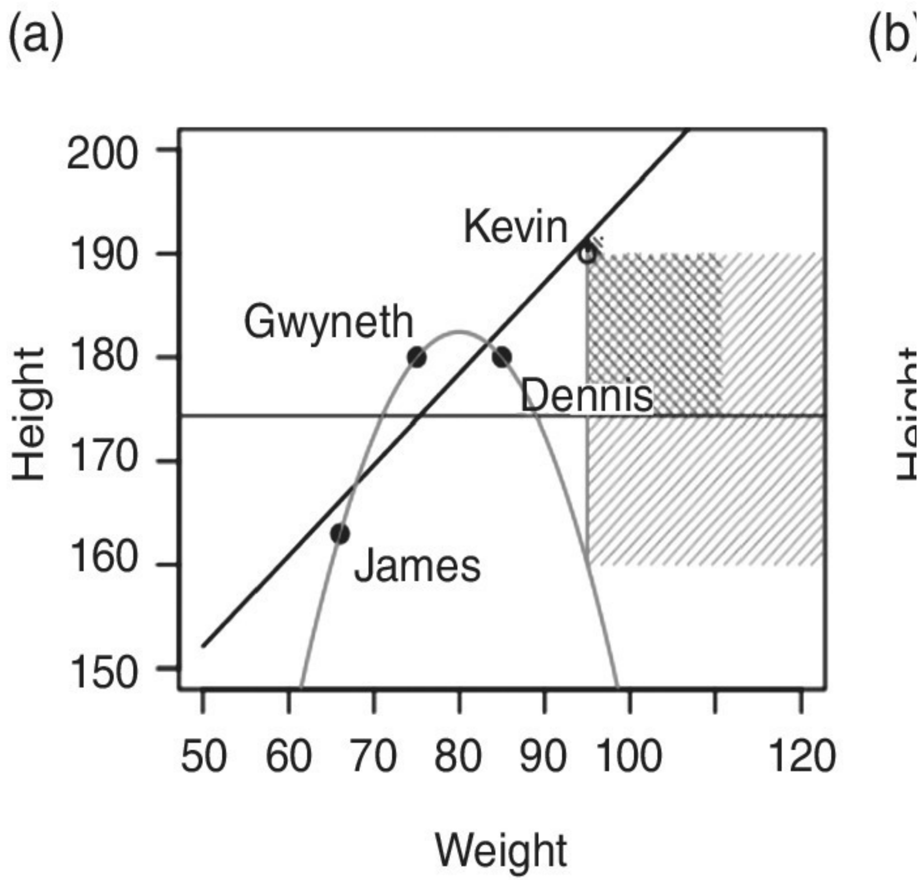
gps

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

- linear regression
- evaluation of regression models
- other algorithms
- bias & variance decomposition of error

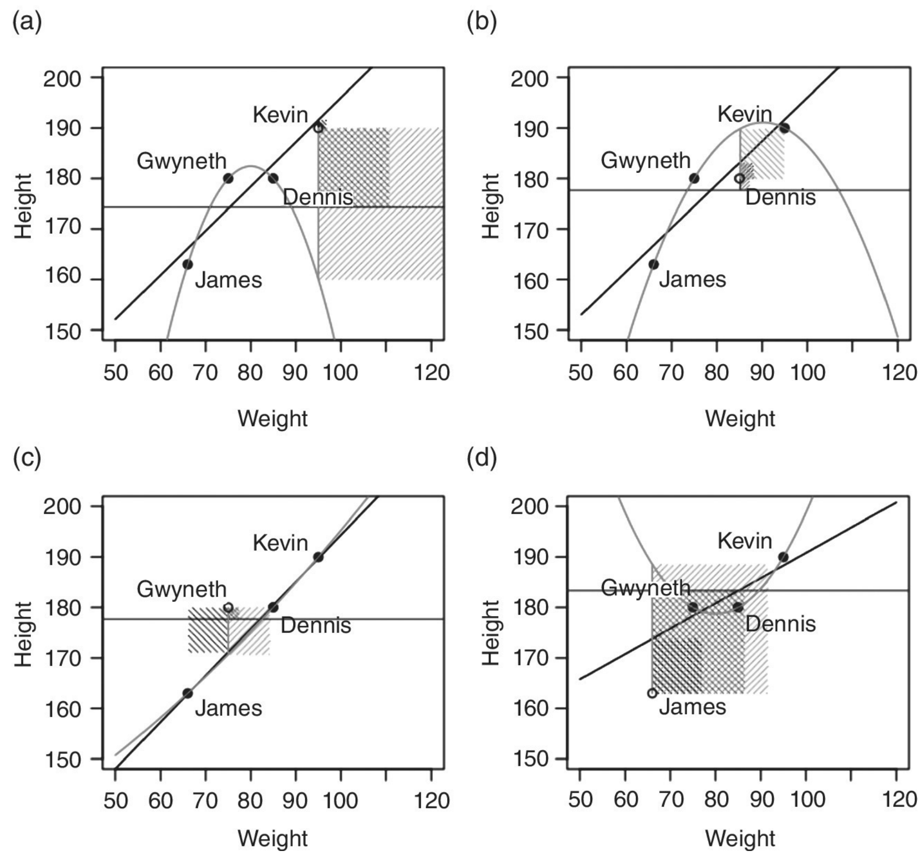
bias

- 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