

Data pre-processing using Weka

Cont..

- Pre-processing of data in weka is done by using filters
- Weka filters can be used to modify datasets in a systematic fashion
- Filters in weka, two types
 - Supervised
 - Unsupervised
- The most commonly used filters in weka are unsupervised filters
- Each filters in both unsupervised and supervised are again classified in to
 - Attribute filters
 - Instance filters

Cont..

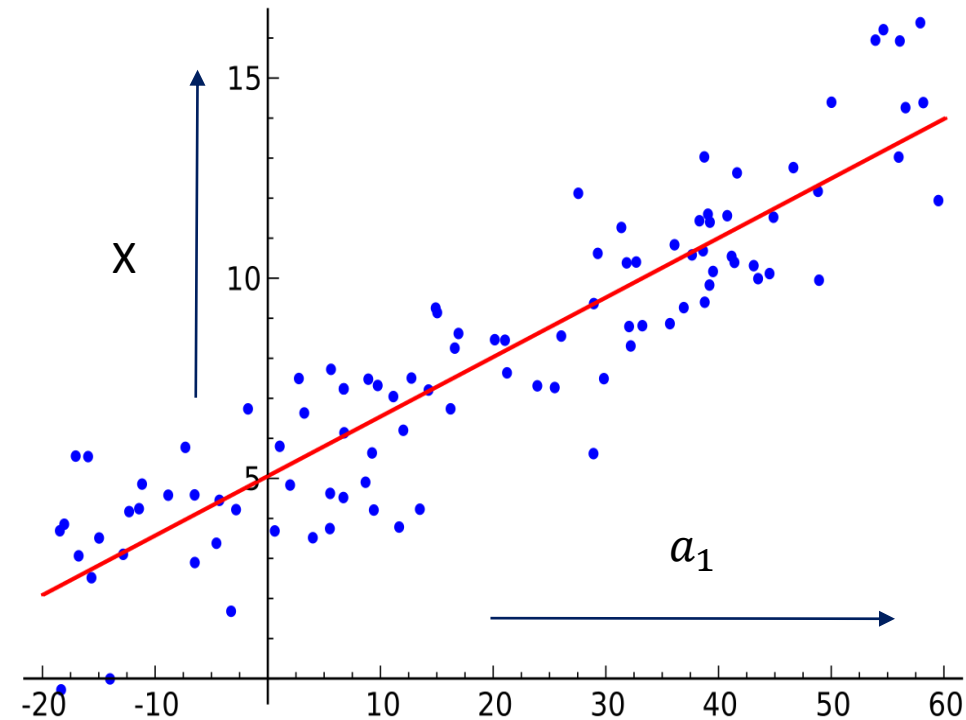
- Some of pre-processing steps used in weka are
 - normalizing
 - renaming the attributes
 - removing an attribute
 - removing the instances
 - randomize
 - remove misclassified

Linear regression using weka

Introduction

- In a regression problem what you are trying to predict is a numerical value
- Linear regression models are used for numeric prediction
- Linear regression is a classical statistical method for predicting numeric classes

Linear regression model



Looking for the best straight line fit

Cont..

$$x = w_0 + w_1 a_1 + w_2 a_2 + w_3 a_3 + \cdots + w_k a_k$$

- Here we are multiplying each attribute values with weights
- The values for weights are calculated using the training data
- Once the weight values are computed, the model can predict the value for each training instance

Cont..

- Predicted value for the first training instance “ $a^{(1)}$ ”

$$w_0 a_0^{(1)} + w_1 a_1^{(1)} + w_2 a_2^{(1)} + \dots + w_k a_k^{(1)} = \sum_{j=0}^k w_j a_j^{(1)}$$

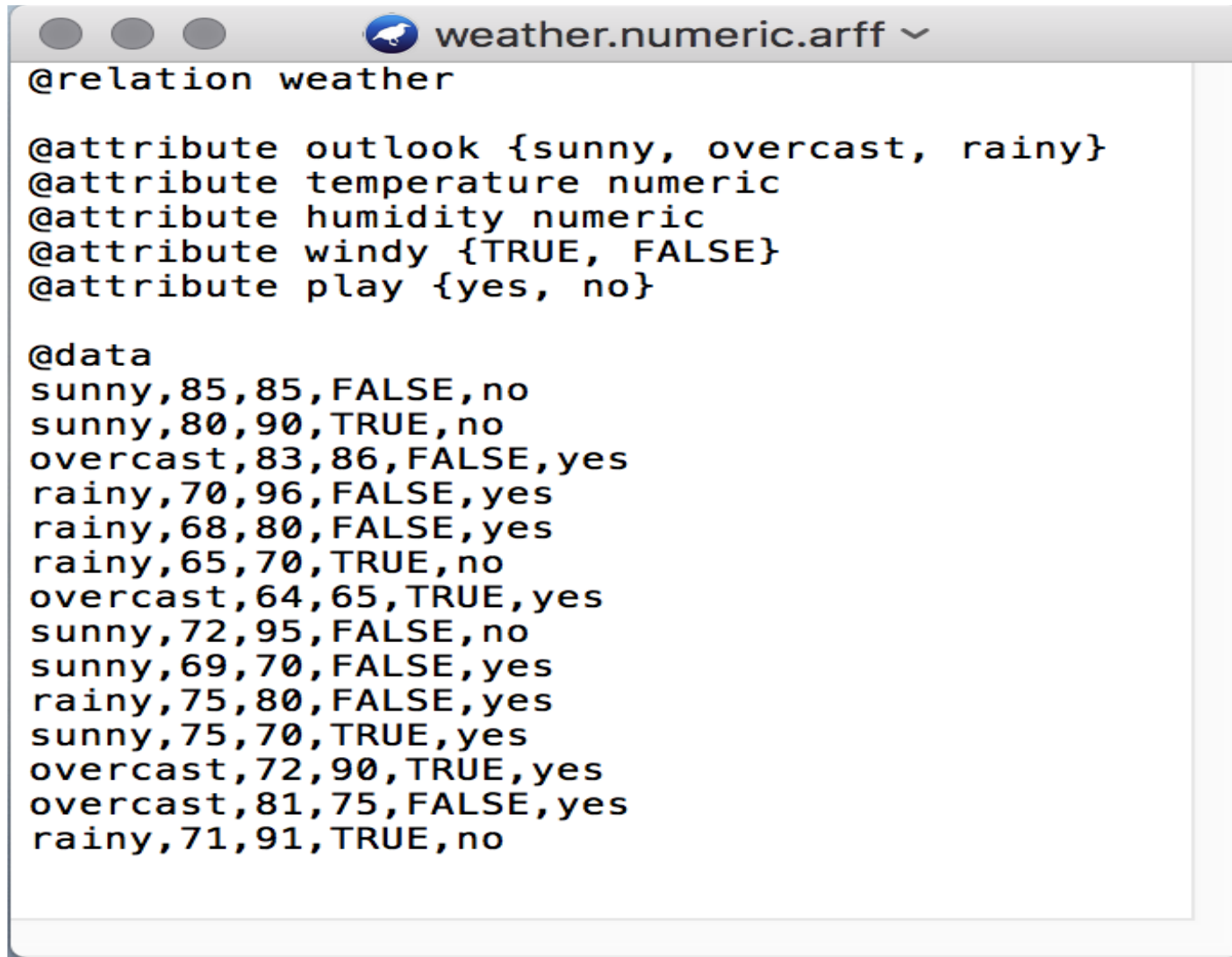
- Choose the weights to minimize squared error on training data

$$\sum_{j=1}^n \left[\underbrace{x^{(i)}}_{\text{Actual Value}} - \underbrace{\sum_{j=0}^k w_j a_j^{(i)}}_{\text{Predicted value}} \right]^2$$

Lab session

- Step 1: load the regression data set using the open file tab
- Step 2: choose the linear regression model (functions › linear regression)
- Step 3: Run the model using start button
- Step 4: Examine the model

ARFF file format



```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```