11 Decision Trees

Solutions

- Exercise 1.1: Decision Tree
  - I(9, 6) = 0.97
- Calculate gain for all the attributes
  - Age:
    - E(Age) = 5/15 * I(2, 3)+ 5/15 * I(3, 2)+ 5/15 * I(4, 1) = 0.88
    - Gain(Age) = 0.97 – 0.88 = 0.09

Age	Has job	Owns house	Credit rating	Approve loan
Young	False	False	Fair	No
Young	False	False	Good	No
Young	True	False	Good	Yes
Young	True	True	Fair	Yes
Young	False	False	Fair	No
Middle	False	False	Fair	No
Middle	False	False	Good	No
Middle	True	True	Good	Yes
Middle	False	True	Excellent	Yes
Middle	False	True	Excellent	Yes
Old	False	True	Excellent	Yes
Old	False	True	Good	Yes
Old	True	False	Good	Yes
Old	True	False	Excellent	Yes
Old	False	False	Fair	No

11 Decision Trees

Solutions

– Has Job:

- E(..) = 10/15 * I(4, 6)+
  5/15 * I(5, 0)
  = 0.64
- Gain(..) = 0.97 – 0.64
  = 0.33

– Owns house:

- E(..) = 9/15 * I(3, 6)+
  6/15 * I(6, 0)
  = 0.54
- Gain(..) = 0.97 – 0.54
  = 0.43

Age	Has job	Owns house	Credit rating	Approve loan
Young	False	False	Fair	No
Young	False	False	Good	No
Young	True	False	Good	Yes
Young	True	True	Fair	Yes
Young	False	False	Fair	No
Middle	False	False	Fair	No
Middle	False	False	Good	No
Middle	True	True	Good	Yes
Middle	False	True	Excellent	Yes
Middle	False	True	Excellent	Yes
Old	False	True	Excellent	Yes
Old	False	True	Good	Yes
Old	True	False	Good	Yes
Old	True	False	Excellent	Yes
Old	False	False	Fair	No

11 Decision Trees

Solutions

– Credit rating:

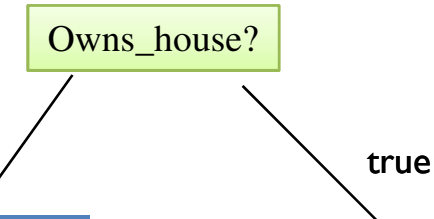
- E(..) = 5/15 * I(1, 4)+
  6/15 * I(4, 2)+
  4/15 * I(4, 0)
  = 0.60
- Gain(..) = 0.97 – 0.60
  = 0.37

Age	Has job	Owns house	Credit rating	Approve loan
Young	False	False	Fair	No
Young	False	False	Good	No
Young	True	False	Good	Yes
Young	True	True	Fair	Yes
Young	False	False	Fair	No
Middle	False	False	Fair	No
Middle	False	False	Good	No
Middle	True	True	Good	Yes
Middle	False	True	Excellent	Yes
Middle	False	True	Excellent	Yes
Old	False	True	Excellent	Yes
Old	False	True	Good	Yes
Old	True	False	Good	Yes
Old	True	False	Excellent	Yes
Old	False	False	Fair	No

11 Decision Trees

Solutions

– Split through Own_house



Age	Has job	Credit rating	Approve loan
Young	False	Fair	No
Young	False	Good	No
Young	True	Good	Yes
Young	False	Fair	No
Middle	False	Fair	No
Middle	False	Good	No
Old	True	Good	Yes
Old	True	Excellent	Yes
Old	False	Fair	No

Age	Has job	Credit rating	Approve loan
Young	True	Fair	Yes
Middle	True	Good	Yes
Middle	False	Excellent	Yes
Middle	False	Excellent	Yes
Old	False	Excellent	Yes
Old	False	Good	Yes

11 Decision Trees

Solutions

– I(3, 6) = 0.91

– Age:

- E(Age) = 4/9 * I(1, 3)+
  2/9 * I(0, 2)+
  3/9 * I(2, 1)
  = 0.66
- Gain(Age) = 0.91 – 0.66
  = 0.25

Owns_house?

false

Age	Has job	Credit rating	Approve loan
Young	False	Fair	No
Young	False	Good	No
Young	True	Good	Yes
Young	False	Fair	No
Middle	False	Fair	No
Middle	False	Good	No
Old	True	Good	Yes
Old	True	Excellent	Yes
Old	False	Fair	No

11 Decision Trees

Solutions

– Has Job:

- E(…) = 6/9 * I(0, 6)+
  3/9 * I(3, 0)
  = 0
- Gain(…) = 0.91
- We can stop and split through Has job

Owns_house?

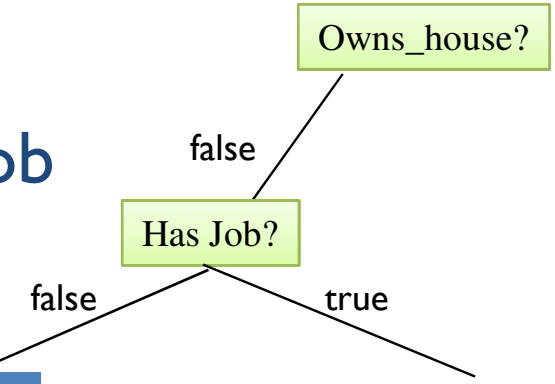
false

Age	Has job	Credit rating	Approve loan
Young	False	Fair	No
Young	False	Good	No
Young	True	Good	Yes
Young	False	Fair	No
Middle	False	Fair	No
Middle	False	Good	No
Old	True	Good	Yes
Old	True	Excellent	Yes
Old	False	Fair	No

11 Decision Trees

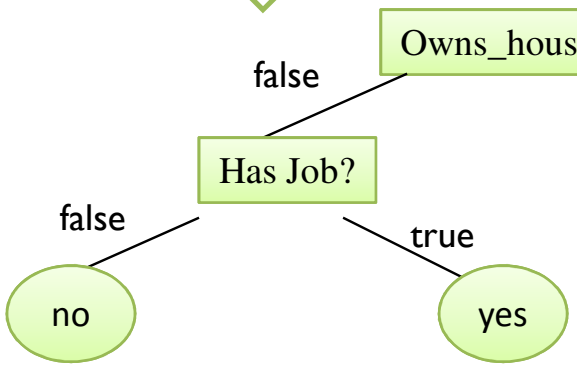
Solutions

– Split through Has job



Age	Has job	Credit rating	Approve loan
Young	False	Fair	No
Young	False	Good	No
Young	False	Fair	No
Middle	False	Fair	No
Middle	False	Good	No
Old	False	Fair	No

Age	Has job	Credit rating	Approve loan
Young	True	Good	Yes
Old	True	Good	Yes
Old	True	Excellent	Yes



Solutions

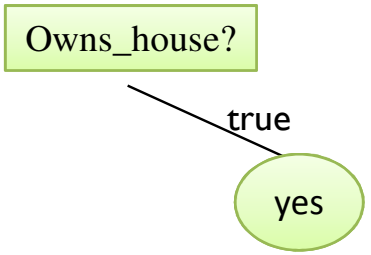
– Recur into the right side

- Stop condition met

Owns_house?

true

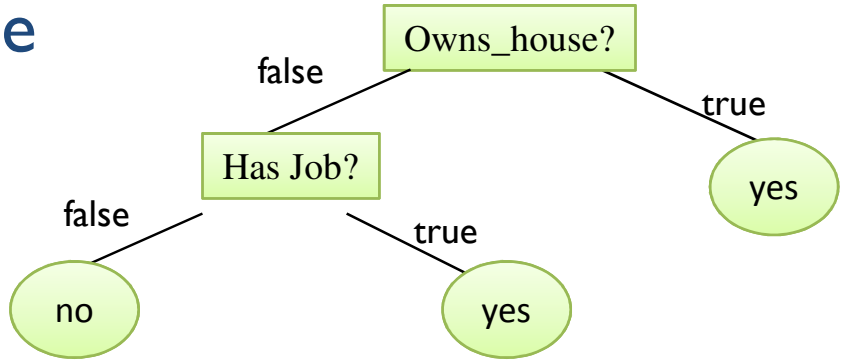
Age	Has job	Credit rating	Approve loan
Young	True	Fair	Yes
Middle	True	Good	Yes
Middle	False	Excellent	Yes
Middle	False	Excellent	Yes
Old	False	Excellent	Yes
Old	False	Good	Yes



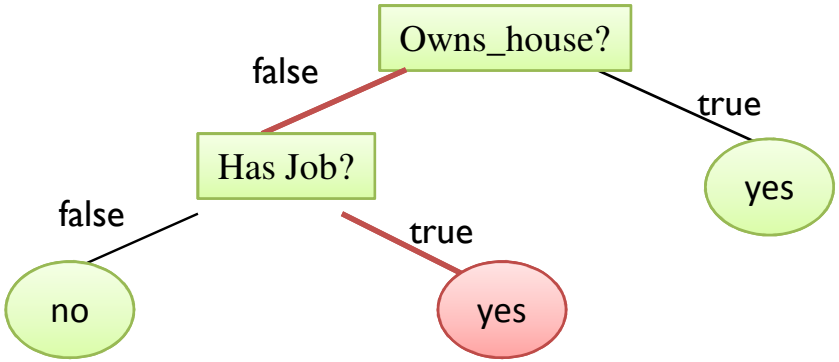
11 Decision Trees

Solutions

– Resulting tree



– X = Senior person with job, doesn't own a house and has good credit rating



11 Decision Trees

Solutions

- Exercise 1.2: Naïve Bayesian Classification
  - P(p) = 9/15
  - P(n) = 6/15

Age attribute	
P(youth|p) = 2/9	P(youth|n) = 3/6
P(middle|p) = 3/9	P(middle|n) = 2/6
P(senior|p) = 4/9	P(senior|n) = 1/6

Has Job	
P(true|p) = 5/9	P(true|n) = 0/6
P(false|p) = 4/9	P(false|n) = 6/6

Owns house	
P(true|p) = 6/9	P(true|n) = 0/6
P(false|p) = 3/9	P(false|n) = 6/6

Age	Has job	Owns house	Credit rating	Approve loan
Young	False	False	Fair	No
Young	False	False	Good	No
Young	True	False	Good	Yes
Young	True	True	Fair	Yes
Young	False	False	Fair	No
Middle	False	False	Fair	No
Middle	False	False	Good	No
Middle	True	True	Good	Yes
Middle	False	True	Excellent	Yes
Middle	False	True	Excellent	Yes
Old	False	True	Excellent	Yes
Old	False	True	Good	Yes
Old	True	False	Good	Yes
Old	True	False	Excellent	Yes
Old	False	False	Fair	No

11 Decision Trees

Solutions

- Exercise 1.2: Naïve Bayesian Classification

Credit rating	
P(fair|p) = 1/9	P(fair|n) = 4/6
P(good|p) = 4/9	P(good|n) = 2/6
P(excellent|p) = 4/9	P(excellent|n) = 0/6

Age	Has job	Owns house	Credit rating	Approve loan
Young	False	False	Fair	No
Young	False	False	Good	No
Young	True	False	Good	Yes
Young	True	True	Fair	Yes
Young	False	False	Fair	No
Middle	False	False	Fair	No
Middle	False	False	Good	No
Middle	True	True	Good	Yes
Middle	False	True	Excellent	Yes
Middle	False	True	Excellent	Yes
Old	False	True	Excellent	Yes
Old	False	True	Good	Yes
Old	True	False	Good	Yes
Old	True	False	Excellent	Yes
Old	False	False	Fair	No

11 Decision Trees

Solutions

- Exercise 1.2: Naïve Bayesian Classification
  - New data:
    - X = Senior person with job, doesn't own a house and has good credit rating
    - P(X|p)*P(p) = 4/9 * 5/9 * 3/9 * 4/9 * 9/15 = 0.021
    - P(X|n)*P(n) = 1/6 * 0 * … = 0
    - **0.02 > 0 So X gets the credit**

Age attribute	
P(youth|p) = 2/9	P(youth|n) = 3/6
P(middle|p) = 3/9	P(middle|n) = 2/6
P(senior|p) = 4/9	P(senior|n) = 1/6

Has Job	
P(true|p) = 5/9	P(true|n) = 0/6
P(false|p) = 4/9	P(false|n) = 6/6

Owns house	
P(true|p) = 6/9	P(true|n) = 0/6
P(false|p) = 3/9	P(false|n) = 6/6

Credit rating	
P(fair|p) = 1/9	P(fair|n) = 4/6
P(good|p) = 4/9	P(good|n) = 2/6
P(excellent|p) = 4/9	P(excellent|n) = 0/6