CS534 Artificial Intelligence. Dept. of Computer Science. WPI
Homework Solutions

Solutions Problems 1 & 2 by Dmitriy Janaliyev and Prof. Ruiz
Solutions Problem 3 by Dmitriy Janaliyev, Robert Martin, and Prof. Ruiz

1. Decision Trees

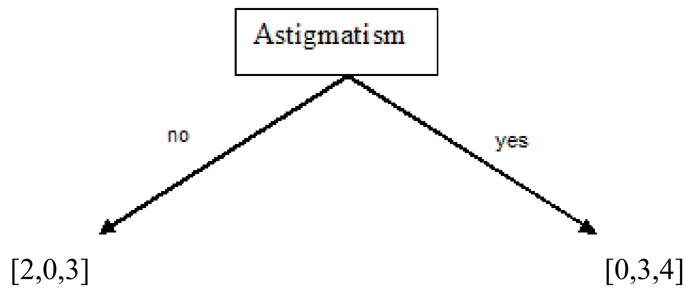
1.1 Model Construction.

Let's compute entropy for each of the attributes over the 12 data instances. Below each attribute value, we write the counts in each class as [soft,hard,none]

Contact lenses	soft	hard	none
Age			
young [1,1,1]	$\frac{3}{12}\left[-\frac{1}{3}\log_2\frac{1}{3}\right.$	$-\frac{1}{3}\log_2\frac{1}{3}$	$\left.-\frac{1}{3}\log_2\frac{1}{3}\right]$ +
pre-presbyopic [1,1,3]	$\frac{5}{12}\left[-\frac{1}{5}\log_2\frac{1}{5}\right.$	$-\frac{1}{5}\log_2\frac{1}{5}$	$\left.-\frac{3}{5}\log_2\frac{3}{5}\right]$ +
presbyopic [0,1,3]	$\frac{4}{12}\left[\ 0\right.$	$-\frac{1}{4}\log_2\frac{1}{4}$	$\left.-\frac{3}{4}\log_2\frac{3}{4}\right]$ =
	$=\frac{3}{12}\left[0.5+0.5+0.5\right]+\frac{5}{12}\left[0.46+0.46+0.42\right]+\frac{4}{12}\left[0.5+0.3\right]=1.2$		
Astigmatism			
no [2,0,3]	$\frac{5}{12}\left[-\frac{2}{5}\log_2\frac{2}{5}\right.$	0	$\left.-\frac{3}{5}\log_2\frac{3}{5}\right]$ +
yes [0,3,4]	$\frac{7}{12}\left[\ 0\right.$	$-\frac{3}{7}\log_2\frac{3}{7}$	$\left.-\frac{4}{7}\log_2\frac{4}{7}\right]$ =
	$=\frac{5}{12}\left[0.52+0.42\right]+\frac{7}{12}\left[0.51+0.46\right]=0.96$		
Tear-prod rate			
reduced [0,0,4]	$\frac{4}{12}\left[\ 0\right.$	0	$\left.-\frac{4}{4}\log_2\frac{4}{4}\right]$ +
normal [2,3,3]	$\frac{8}{12}\left[-\frac{2}{8}\log_2\frac{2}{8}\right.$	$-\frac{3}{8}\log_2\frac{3}{8}$	$\left.-\frac{3}{8}\log_2\frac{3}{8}\right]$ =
	$=0+\frac{8}{12}\left[0.5+0.525+0.525\right]=1.03$		

The attribute Astigmatism has the lowest entropy, so we choose it as the root of the decision tree.

Since each of the two resulting children are heterogeneous, we need to decide which attribute to choose to split the data instances in each of the two branches (Astigmatism = no; Astigmatism = yes).



For Astigmatism = no:

Contact lenses	soft	hard	none	
Age				
young [1,0,0]	$\frac{1}{5}\left[-\frac{1}{1}\log_2\frac{1}{1}\right.$	0	$\left.0\right]$	+
pre-presbyopic [1,0,1]	$\frac{2}{5}\left[-\frac{1}{2}\log_2\frac{1}{2}\right.$	0	$\left.-\frac{1}{2}\log_2\frac{1}{2}\right]$	+
presbyopic [0,0,2]	$\frac{2}{5}\left[\ 0\right.$	0	$\left.-\frac{2}{2}\log_2\frac{2}{2}\right]$	=
	$= 0+\frac{2}{5}[0.5+0.5]+0=0.4$			
Tear-prod rate				
reduced [0,0,2]	$\frac{2}{5}\left[\ 0\right.$	0	$\left.-\frac{2}{2}\log_2\frac{2}{2}\right]$	+
normal [2,0,1]	$\frac{3}{5}\left[-\frac{2}{3}\log_2\frac{2}{3}\right.$	0	$\left.-\frac{1}{3}\log_2\frac{1}{3}\right]$	=
	$=0+\frac{3}{5}[0.39+0+0.5]=0.53$			

Hence, for the branch with Astigmatism = no, the chosen attribute is Age, as it is the one with the lowest entropy.

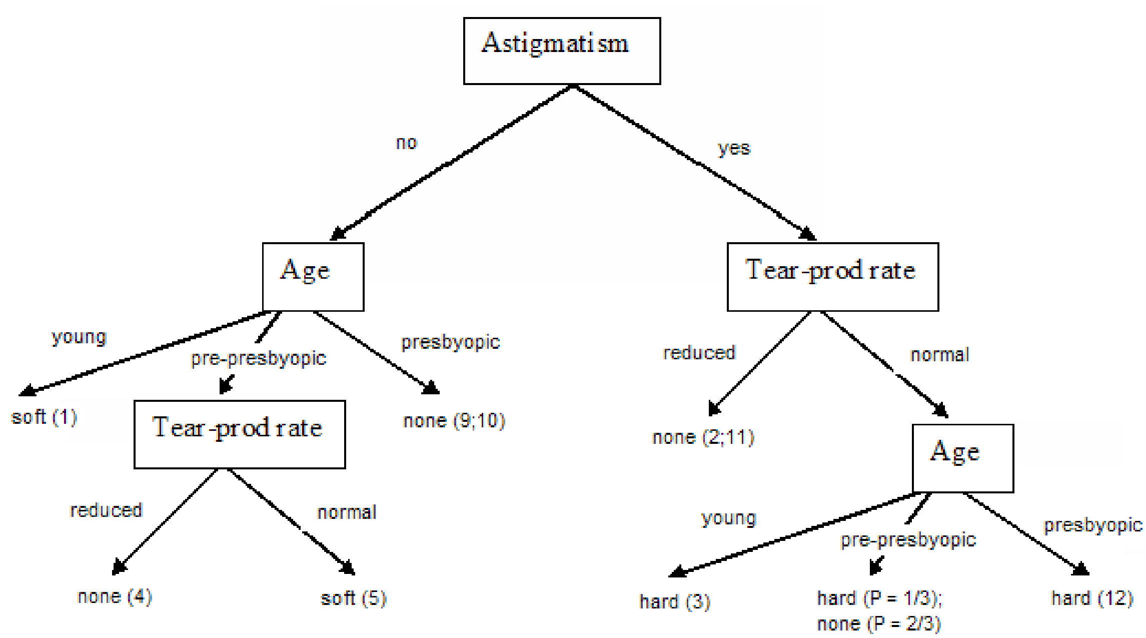
For Astigmatism = yes:

Contact lenses	soft	hard	none	
Age				
young [0,1,1]	$\frac{2}{7}[$ 0	$-\frac{1}{2}\log_2\frac{1}{2}$	$-\frac{1}{2}\log_2\frac{1}{2}]$	+
pre-presbyopic [0,1,2]	$\frac{3}{7}[$ 0	$-\frac{1}{3}\log_2\frac{1}{3}$	$-\frac{2}{3}\log_2\frac{2}{3}]$	+
presbyopic [0,1,1]	$\frac{2}{7}[$ 0	$-\frac{1}{2}\log_2\frac{1}{2}$	$-\frac{1}{2}\log_2\frac{1}{2}]$	=
	$= \frac{2}{7}[0+0.5+0.5] + \frac{3}{7}[0+0.53+0.39] + \frac{2}{7}[0+0.5+0.5] = \frac{6.76}{7} = 0.96$			
Tear-prod rate				
reduced [0,0,2]	$\frac{2}{7}[$ 0	0	$-\frac{2}{2}\log_2\frac{2}{2}]$	+
normal [0,3,2]	$\frac{5}{7}[$ 0	$-\frac{3}{5}\log_2\frac{3}{5}$	$-\frac{2}{5}\log_2\frac{2}{5}]$	=
	$= 0 + \frac{5}{7}[0+0.42+0.52] = 0.67$			

For the second branch (Astigmatism = yes), the chosen attribute is Tear-prod rate as it has the lowest entropy.

Now we need to continue splitting the resulting children that are still heterogeneous. Since only one attribute remains as an option in each of the branches, we use that remaining attribute (no need to do entropy calculations as we don’t need to choose among several options).

The final decision tree is depicted on the next page (numbers inside parentheses on the leaf nodes denote the “ids” of the dataset instances that belong to that leaf).



Note that the node at the end of the branch (Astigmatism = yes) → (Tear-prod rate = normal) → (Age = pre-presbyopic) is still heterogeneous but we don't have more attributes to split that node. We consider two options for the prediction value assigned to that leaf: (1) to return probabilities of the two outcomes (classes), based on the number of corresponding dataset instances for each of them (as shown on the tree above); or (2) to select the majority class for that node, namely Contact-lenses=none (as there are 2 "none" data instances vs. 1 "hard" data instance in that leaf).

1.2. Model deployment.
For the test instances given, we have the following predictions:

test1: "young, no, reduced, none".	The constructed decision tree predicts: _soft_
test2: "pre-pre, yes, reduced, none".	The constructed decision tree predicts: _none
test3: "presbyopic, no, normal, soft".	The constructed decision tree predicts: _none_
test4: "presbyopic, yes, normal, hard".	The constructed decision tree predicts: _hard_

For this given test set, the accuracy is 50% (2 correct predictions and 2 incorrect predictions).

2. **Naïve Bayes Models**

2.1. **Model Construction.** Let's compute prior and conditional probabilities:

Contact-lenses:	**soft**	**hard**	**none**
probability:	(5 + 1)/27	(4 + 1)/27	(15 + 1)/27
Age:			
young	(2 + 1)/8	(2 + 1)/7	(4 + 1)/18
pre-presbyopic	(2 + 1)/8	(1 + 1)/7	(5 + 1)/18
presbyopic	(1 + 1)/8	(1 + 1)/7	(6 + 1)/18
Spectacle-prescription			
myope	(2 + 1)/7	(3 + 1)/6	(7 + 1)/17
hypermetrope	(3 + 1)/7	(1 + 1)/6	(8 + 1)/17
Astigmatism			
no	(5 + 1)/7	(0 + 1)/6	(7 + 1)/17
yes	(0 + 1)/7	(4 + 1)/6	(8 + 1)/17
Tear-prod-rate			
reduced	(0 + 1)/7	(0 + 1)/6	(12 + 1)/17
normal	(5 + 1)/7	(4 + 1)/6	(3 + 1)/17

2.1. **Model Deployment.** Now, let's use the constructed Naive Bayes model, to predict results for the given data instance: "presbyopic, hypermetrope, yes, normal"

predicted v = $\arg\max_{v}$ P(Contact-lenses = v)*P(Age = presbyopic | Contact-lenses = v)*
P(Spectacle-prescription = hypermetrope | Contact-lenses = v)*
P(Astigmatism = yes | Contact-lenses = v)*
P(Tear-prod-rate = normal | Contact-lenses = v)

v = soft: P(Contact-lenses = soft)*P(Age = presbyopic | Contact-lenses = soft)*
P(Spectacle-prescription = hypermetrope | Contact-lenses = soft)*
P(Astigmatism = yes | Contact-lenses = soft)*
P(Tear-prod-rate = normal | Contact-lenses = soft) =
= (6/27)*(2/8)*(4/7)*(1/7)*(6/7) = 288/74088 = 0.00389

v = hard: P(Contact-lenses = hard)*P(Age = presbyopic | Contact-lenses = hard)*
P(Spectacle-prescription = hypermetrope | Contact-lenses = hard)*
P(Astigmatism = yes | Contact-lenses = hard)*
P(Tear-prod-rate = normal | Contact-lenses = hard) =
= (5/27)*(2/7)*(2/6)*(5/6)*(5/6) = 500/40824 = 0.01225

v = none: P(Contact-lenses = none)*P(Age = presbyopic | Contact-lenses = none)*
P(Spectacle-prescription = hypermetrope | Contact-lenses = none)*
P(Astigmatism = yes | Contact-lenses = none)*
P(Tear-prod-rate = normal | Contact-lenses = none) =
= (16/27)*(7/18)*(9/17)*(9/17)*(4/17) = 36288/2387718 = 0.01519

Hence the Naïve Bayes classifier predicts "none" for this data instance.

3. **Bayesian Networks**

3.1. **Model Construction.** Let's construct prior and conditional probability tables.

Contact-lenses	soft	hard	none
	(5 + 1)/27	(4 + 1)/27	(15 + 1)/27

Age / **Contact-lenses**	young	pre-presbyopic	presbyopic
soft	(2 + 1)/8	(2 + 1)/8	(1 + 1)/8
hard	(2 + 1)/7	(1 + 1)/7	(1 + 1)/7
none	(4 + 1)/18	(5 + 1)/18	(6 + 1)/18

Spectacle-prescription / **Contact-lenses**	myope	hypermetrope
soft	(2 + 1)/7	(3 + 1)/7
hard	(3 + 1)/6	(1 + 1)/6
none	(7 + 1)/17	(8 + 1)/17

Tear-prod-rate / **Contact-lenses**	reduced	normal
soft	(0 + 1)/7	(5 + 1)/7
hard	(0 + 1)/6	(4 + 1)/6
none	(12 + 1)/17	(3 + 1)/17

Age	**Spectacle-prescription** / **Astigmatism**	no	yes
young	myope	(2 + 1)/6	(2 + 1)/6
young	hypermetrope	(2 + 1)/6	(2 + 1)/6
pre-presbyopic	myope	(2 + 1)/6	(2 + 1)/6
pre-presbyopic	hypermetrope	(2 + 1)/6	(2 + 1)/6
presbyopic	myope	(2 + 1)/6	(2 + 1)/6
presbyopic	hypermetrope	(2 + 1)/6	(2 + 1)/6

3.2. **Model Deployment.** Given the test data instance "yes, normal":

predicted classification =

$= \arg\max_v$ P(Contact-lenses = v | Astigmatism = yes & Tear-prod-rate = normal)

$$= \arg\max_v \frac{P(Astigmatism = yes \,\&\, Tear-prod-rate = normal \mid Contact-lenses = v) * P(Contact-lenses = v)}{P(Astigmatism = yes \,\&\, Tear-prod-rate = normal)}$$

since the denominator is constant for any value v of Contact-lenses, we can just ignore it:

$= \arg\max_v$ P(Astigmatism = yes & Tear-prod-rate = normal | Contact-lenses = v)*P(Contact-lenses = v)

$= \arg\max_v$ P(Astigmatism = yes & Tear-prod-rate = normal | **Age & Spec-Presc** & Contact-lenses = v**)***

*P(**Age & Spec-Presc**| Contact-lenses = v)
*P(Contact-lenses = v)

since Astigmatism and Tear-prod-rate are independent of each other given their parents (namely Age, Spect-presc, and Contact-lenses) then:

$= \arg\max_v$ P(Astigmatism = yes | **Age & Spec-Presc** & Contact-lenses = v)

*P(Tear-prod-rate = normal | **Age & Spec-Presc** & Contact-lenses = v)
*P(**Age & Spec-Presc**| Contact-lenses = v)
*P(Contact-lenses = v)

since Astigmatism is independent of Contact-lenses given Age & Spec-Presc; Tear-prod-rate is independent of Age & Spec-Presc given Contact-lenses; and Age and Spec-Presc are independent of each other given Contact-lenses, then:

$= \arg\max_v$ P(Astigmatism = yes | **Age & Spec-Presc**)

* P(Tear-prod-rate = normal | Contact-lenses = v)
* P(**Age**| Contact-lenses = v)
* P(**Spec-Presc**| Contact-lenses = v)
* P(Contact-lenses = v)

reorganizing the terms for convenience, and abbreviating names to save space:

$= \arg\max_v$ P(Contact-lenses = v)

* P(Tear-prod-rate = normal | Contact-lenses = v)
*P(Astigmatism = yes | **Age & Spec-Presc**)
* P(**Age**| Contact-lenses = v)
* P(**Spec-Presc**| Contact-lenses = v)

$= \arg\max_v$ P(Contact-lenses = v) *

* P(Tear-prod-rate = normal | Contact-lenses = v) *
* [P(Astig=yes | Age=young & SP=myope)* P(Age=young| CL=v)* P(SP=myope| CL=v) +
+ P(Astig=yes | Age=young & SP=hyperm)*P(Age=young| CL=v)* P(SP=hypermyope| CL=v) +
+ P(Astig=yes | Age=pre-p & SP=myope)*P(Age=pre-p| CL=v)*P(SP=myope| CL=v) +
+ P(Astig=yes | Age=pre-p & SP=hyperm)*P(Age=pre-p| CL=v)*P(SP=hyperm| CL=v) +
+ P(Astig= yes | Age=presb & SP=myope)*P(Age=presb| CL=v)*P(SP=myope| CL=v) +
+ P(Astig = yes | Age=presbyopic & SP=hyperm)*P(Age=presb| CL=v)*P(SP=hyperm| CL=v)
]

$= \arg\max_{v}$ P(Contact-lenses = v) *
* P(Tear-prod-rate = normal | Contact-lenses = v) *
* (3/6) [P(Age=young| CL=v)* P(SP=myope| CL=v) +
+ P(Age=young| CL=v)* P(SP=hypermyope| CL=v) +
+ P(Age=pre-p| CL=v)*P(SP=myope| CL=v) +
+ P(Age=pre-p| CL=v)*P(SP=hyperm| CL=v) +
+ P(Age=presb| CL=v)*P(SP=myope| CL=v) +
+ P(Age=presb| CL=v)*P(SP=hyperm| CL=v)
]

$= \arg\max_{v}$ P(Contact-lenses = v) *
* P(Tear-prod-rate = normal | Contact-lenses = v) *
* (3/6) [P(Age=young| CL=v)* [P(SP=myope| CL=v) + P(SP=hypermyope| CL=v)] +
+ P(Age=pre-p| CL=v)*[P(SP=myope| CL=v) + P(SP=hyperm| CL=v)] +
+ P(Age=presb| CL=v)*[P(SP=myope| CL=v) + P(SP=hyperm| CL=v)]
]

$= \arg\max_{v}$ P(Contact-lenses = v) *
* P(Tear-prod-rate = normal | Contact-lenses = v) *
* (3/6) [P(Age=young| CL=v)* [1] + P(Age=pre-p| CL=v)*[1] + P(Age=presb| CL=v)*[1]]

$= \arg\max_{v}$ P(Contact-lenses = v) *
* P(Tear-prod-rate = normal | Contact-lenses = v) *
* (3/6) [1]

Note: *The calculations above could have been simplified from the beginning by noticing that P(Astigmatism=yes| its parents) = 3/6 regardless of the particular values taken by Astigmatism's parents Age and Spectacle-prescription.*

Let's consider now each possible value v of Contact-lenses:

v = soft: P(Contact-lenses =soft)*P(Tear-prod-rate=normal | Contact-lenses=soft)*(3/6)
= (6/27)*(6/7)*(3/6) = 18/189= 0.0952

v = hard: P(Contact-lenses =hard)*P(Tear-prod-rate=normal | Contact-lenses=hard)*(3/6)
= (5/27)*(5/6)*(3/6) = 25/324= 0.0772

v = none: P(Contact-lenses =none)*P(Tear-prod-rate=normal | Contact-lenses=none)*(3/6)
= (16/27)*(4/17)*(3/6) = 192/2754= 0.0697

Thus, the predicted value for this test data instance is Contact-lenses = soft, since v=soft produces the highest probability.