**CSci 384: Artificial Intelligence** **Spring, 2020**

**Instructor: Dr. M. E. Kim** **Date: April 23, 2020**

# Assignment 7: Solution

## Q1. [100] Inductive Decision Tree Learning

The table below shows training and test examples typical of a classification task.

| *Example* | *Author* | *Thread* | *Length* | *WhereRead* | *UserAction* |
|---|---|---|---|---|---|
| $e_1$ | known | new | long | home | skips |
| $e_2$ | unknown | new | short | work | reads |
| $e_3$ | unknown | follow Up | long | work | skips |
| $e_4$ | known | follow Up | long | home | skips |
| $e_5$ | known | new | short | home | reads |
| $e_6$ | known | follow Up | long | work | skips |
| $e_7$ | unknown | follow Up | short | work | skips |
| $e_8$ | unknown | new | short | work | reads |
| $e_9$ | known | follow Up | long | home | skips |
| $e_{10}$ | known | new | long | work | skips |
| $e_{11}$ | unknown | follow Up | short | home | skips |
| $e_{12}$ | known | new | long | work | skips |
| $e_{13}$ | known | follow Up | short | home | reads |
| $e_{14}$ | known | new | short | work | reads |
| $e_{15}$ | known | new | short | home | reads |
| $e_{16}$ | known | follow Up | short | work | reads |
| $e_{17}$ | known | new | short | home | reads |
| $e_{18}$ | unknown | new | short | work | reads |
| $e_{19}$ | unknown | new | long | work | ? |
| $e_{20}$ | unknown | follow Up | long | home | ? |

The aim is to predict whether a person reads an article posted to a bulletin board given properties of the article.
The input features are *Author*, *Thread*, *Length*, and *WhereRead*. There is one target feature, *UserAction*.
The domain of *Author* is {known, unknown}, the domain of *Thread* is {new, followup}, and so on.
There are **18 *training examples***, each of which has a value for all of the features.
In this data set, *val*(*e*11, *Author*)=*unknown*, *val*(*e*11, *Thread*)=*followUp*, and *val*(*e*11, *UserAction*)=*skips*.
There are **2 *test examples***, **e19** and **e20**, where the user action is unknown.
The aim is to predict the user action for a new example given its values for the input features.

1) [10] Compute the ***initial entropy*** of *UserAction.*

   *For UserAction, reads = skips =* 9, out of 18 examples.

   $$B\left(<\frac{9}{18},\frac{9}{18}>\right) = -(\frac{9}{18}\log_2\frac{9}{18} + \left(1-\frac{9}{18}\right)\log_2(1-\frac{9}{18})) = 1\ bits$$

2) [10] From the training examples ($e_1$ – $e_{18}$), construct an optimal decision tree that classifies the data the best. Show the proper computation of information gain with the entropy of each variable.

If there is conflicting description in an example, give the estimated probabilities of each classification using the relative frequencies.

Initial Entropy of UserAction = 1 from 1).

1.Calculate information gain of each variable:

$$Gain(A) = B\left(\frac{p}{p+n}\right) - \sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B\left(\frac{p_k}{p_k+n_k}\right)$$

- Author: 12 known: 6 skips and 6 reads, 6 unknown: 3 skips and 3 reads

$$Gain(Author) = B\left(\frac{9}{9+9}\right) - \left[\frac{6+6}{9+9} B\left(\frac{6}{6+6}\right) + \frac{3+3}{9+9} B\left(\frac{3}{3+3}\right)\right] = 0\ bits$$

- Thread: 10 new: 3 skips and 7 reads, 8 follow up: 6 skips and 2 reads

$$Gain(Thread) = B\left(\frac{9}{9+9}\right) - \left[\frac{7+3}{9+9} B\left(\frac{7}{7+3}\right) + \frac{2+6}{9+9} B\left(\frac{2}{2+6}\right)\right] \cong .15\ bits$$

- Length: 7 long: 7 skips and 0 reads, 11 short: 2 skips and 9 reads

$$Gain(Length) = B\left(\frac{9}{9+9}\right) - \left[\frac{0+7}{9+9} B\left(\frac{0}{0+7}\right) + \frac{9+2}{9+9} B\left(\frac{9}{9+2}\right)\right] \cong .582\ bits$$

- WhereRead: 8 home: 4 skips and 4 reads, 10 work: 5 skips and 5 reads

$$Gain(whereRead) = B\left(\frac{9}{9+9}\right) - \left[\frac{4+4}{9+9} B\left(\frac{4}{4+4}\right) + \frac{5+5}{9+9} B\left(\frac{5}{5+5}\right)\right] = 0\ bits$$

So, the root of the tree is **Length** whose information gain is the highest.

2. For a subset of data where Length = short, {e2, e5, e7, e8, e11, e13, e14, e15, e16, e17, e18}

$$Gain(A) = B\left(\frac{p}{p+n}\right) - \sum_{k=1}^{d} \frac{p_k+n_k}{p+n} B\left(\frac{p_k}{p_k+n_k}\right)$$

- Author: 6 known: 0 skips and 6 reads, 5 unknown: 2 skips and 3 reads

$$Gain(Author) = B\left(\frac{9}{9+2}\right) - \left[\frac{3+2}{9+2} B\left(\frac{2}{2+2}\right) + \frac{6+0}{9+2} B\left(\frac{6}{6+0}\right)\right] \cong .24\ bits$$

- Thread: 7 new: 0 skips and 7 reads, 4 follow up: 2 skips and 2 reads

$$Gain(Thread) = B\left(\frac{9}{9+2}\right) - \left[\frac{2+2}{9+2} B\left(\frac{2}{2+2}\right) + \frac{7+0}{9+2} B\left(\frac{7}{7+0}\right)\right] \cong .320\ bits$$

- WhereRead: 5 home: 1 skips and 4 reads, 6 work: 1 skips and 5 reads

$$Gain(whereRead) = B\left(\frac{9}{9+2}\right) - \left[\frac{4+1}{9+2} B\left(\frac{4}{4+1}\right) + \frac{5+1}{9+2} B\left(\frac{5}{5+1}\right)\right] \cong .001\ bits$$

So, WhereRead is a root of a subtree of [Length = short]

3. For a subset of data where Thread = followUp, {e2, e5, e8, e14, e15, e17, e18 }

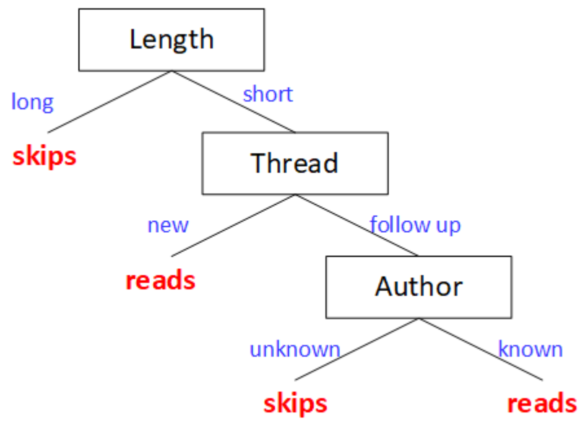
- Author: 2 known: 0 skips and 2 reads, 2 unknown: 2 skips and 0 reads

  $$G(Author) = B\left(\frac{2}{2+2}\right) - \left[\frac{0+2}{2+2}B\left(\frac{0}{0+2}\right) + \frac{2+0}{2+2}B\left(\frac{2}{2+0}\right)\right] = 1\ bits$$

- whereRead: 2 home: 1 skips and 1 reads, 2 work: 1 skips and 1 reads

  $$G(whereRead) = B\left(\frac{2}{2+2}\right) - \left[\frac{1+1}{2+2}B\left(\frac{1}{1+1}\right) + \frac{1+1}{2+2}B\left(\frac{1}{1+1}\right)\right] = 0\ bits$$

  So, Author is a root of a subtree of [Thread=followUp ], which clearly classifies data.




3) [10] Express the hypothesis generated at 2) in the logical expression.

   ∀x UserAction(x) = reads ↔ (Length(x)=Short ∧ Thread(x) = New) ∨ (Length(x) = Short ∧ Thread(x) = Follow UP ∧ Author(x) = Known) where x is a datum

4) [10] For the test data e19 and e20, predict the user's action based on the hypothesis in 2) – 3), respectively.

   Read = (Length=short ^ (Thread=new v Author=known))
   e19: Read = (0 ^ 1) = 0: User will skip
   e20: Read = (0 ^ 0) = 0: User will skip

   From the same training examples (e1 – e18),

5) [30] Decide the following probabilities that can be derived from the given data of the training examples in the table.
   a) [4] P( *UserAction* = reads) = = 9/18 = 1/2
   b) [3] P( *Author* = known | *UserAction* = reads) = 6/9 = 2/3
   c) [3] P( *Author* = known | *UserAction* = skips) = 6/9 = 2/3
   d) [3] P( *Thread* = new | *UserAction* = reads) = 7/9
   e) [3] P( *Thread* = new | *UserAction* = skips) = 3/9
   f) [3] P( *Length* = long | *UserAction* = reads) = 0/9 = 0
   g) [3] P( *Length* = long | *UserAction* = skips) = 7/9
   h) [4] P( *WhereRead* = home | *UserAction* = reads) = 4/9
   i) [4] P( *WhereRead* = home | *UserAction* = skips) = 4/9

6) [10] By means of ***Naïve Bayes Classifier model***, predict the classification, *UserAction*, of the test data **e19**. You have to show the computations of classification probabilities. i.e. decide $C_{NB}$.

e19 <unknown, new, long, work, ?>
P(UserAction = read) = 9/18 = .5 P(UserAction = skip) = 9/18 = .5
P(UserAction = read | e19)
=α P(read) * P(unknown | read) * P(new | read) * P(long | read) * P(work | read)
= α 9/18 * 3/9 * 7/9 * 0/9 * 5/9 = 0
P(UserAction = skips | e19)
= α P(skip) * P(unknown | skip) * P(new | skip) * P(long | skip) * P(work | skip)
= α 9/18 * 3/9 * 3/9 * 7/9 * 5/9 = .5 * .33 * .33 * .78 * .56 = .024α = 1
Since P(UserAction = read | e19) < P(UserAction = skips | e19),
**the Naïve Bayes classifier ($C_{NB}$ ) for e19 is: $C_{NB} = (userAction = skips)$.**

7) [10] What is the predicted classification probability of **e19**, i.e. P($C_{NB}$ | e19), and its prediction?
< P(reads | e19), P(skips | e19) > = α <0, .0265> = <0, 1>, so, P($C_{NB}$ | e19) = 1.

8) [10] What is the predicted classification probability of **e20**, i.e. P($C_{NB}$ | e20), and its prediction?

e20 <unknown, follow-up, long, home>
P(reads| e20)
= P(read) * P(unknown | read) * P(follow-up | read) * P(long | read) * P(home | read)
= 9/18 * 3/9 * 2/9 * 0/9 * 4/9 = 0
P(skips | e20)
= P(skip) * P(unknown | skip) * P(follow-up | skip) * P(long | skip) * P(home | skip)
= 9/18 * 3/9 * 6/9 * 7/9 * 4/9 = .5 * .33 * .67 * .78 * .44 = .0379
Thus $C_{NB}$ = (UserAction=skip)

P(UserAction=skip | e20> = .0379 / (.0379+0) = 1.0

## Q2. [20] Maximum Likelihood (ML) Learning

Suppose we toss a thumbtack 4 times and we observe the sequence [heads, tails, heads, heads].
Let $T_1$, $T_2$, $T_3$ and $T_4$ be random variables such that the value of $T_i$ is the outcome on the $i^{th}$ toss.

1) [10] Give the maximum likelihood (ML) estimate of the probability of heads, P($T_i$ = head).
Hint: Define P(heads) = *p*. Then, compute the parameter value of *p* using ML.

3 heads, 1 tails. P(heads) = p, P(tails) = (1-p)
$L(p) = p^3(1-p)^1$
Log $L(p) = 3\ \log(p) + 1\log(1-p)$
$\frac{d}{dp}\log L(p) = \frac{3}{p} - \frac{1}{1-p} = 0$
$\frac{3}{p} = \frac{1}{1-p}\ solves\ to\ p = \frac{3}{4}$
P($T_i$ = head) = ¾ = .75

2) [10] Suppose that the data is identically independently distributed.
What is the ML estimate of P( $T_1$ = head, $T_2$ = tail, $T_3$ = head, $T_4$ = head) ?

Using the estimated probability P($T_i$ = head) in 1),

P( $T_1$ = head, $T_2$ = tail, $T_3$ = head, $T_4$ = head)
= P(heads) * P(tails) * P(heads) * P(heads)
= .75 * .25 * .75 * .75 = .1055

**Q3. [60] Maximum Likelihood (ML) Learning**

The table shows the examples of SPAM and those of HAM messages which are consisted of some words whose dictionary size is twelve. Suppose that you've received SPAM messages for the 1st 3 days, then HAM messages for the next 5 days, i.e. a data as a sequence of message is <Spam, Spam, Spam, Ham, Ham, Ham, Ham, Ham >.

| ***SPAM*** | ***HAM*** |
|---|---|
| offer is secret | play golf tomorrow |
| click secret link | went play golf |
| secret golf link | secret golf event |
| | golf is tomorrow |
| | golf costs money |

(1) [10] Compute the ***maximum likelihood*** of *SPAM, i.e.* P(*SPAM*)=θ, using a *log-likelihood.*

In data, 3 SPAM messages and 5 HAM messages.
Let P(SPAM) = θ to be estimated. Then, P(HAM) = 1 - θ.
$P(d|h_\theta) = \theta^s(1-\theta)^h$ where s and h are the # of SPAM or HAM messages, respectively.
$L(d|h_\theta) = s{\cdot}log\theta + h{\cdot}log(1-\theta) = 3log\theta + 5\log(1-\theta)$

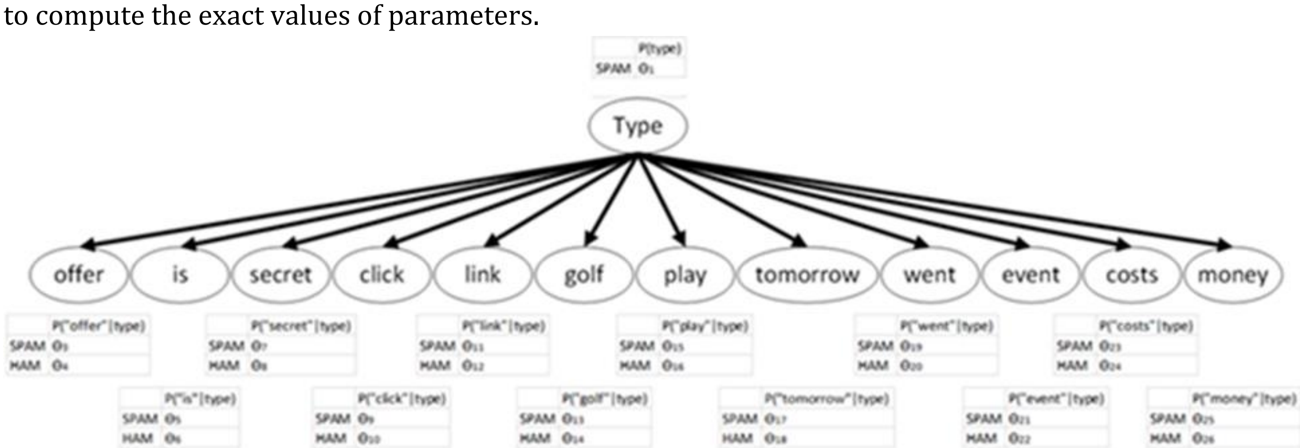
$$\frac{dL(d|h_\theta)}{d\theta} = \frac{3}{\theta} - \frac{5}{1-\theta} = 0 \Rightarrow \theta = \frac{3}{3+5} = \frac{3}{8} = 0.375$$

(2) [10] In the Bayes net of this ML parameter learning,

(a) how many parameters are required?

1 + 2 * 12 = 25 parameters.

(b) Draw the BN with the CPT of the required parameters (e.g. $\theta_1$, $\theta_2$, $\theta_3$, …..). -- You don't yet have to compute the exact values of parameters.



(3) [10] By ML-learning, compute a *parameter value*, P("*secret*" | *SPAM*) and P("*secret*" | *HAM*), respectively, using a log-likelihood.

$$P(W = secret|C = SPAM) = \frac{3}{9} \approx 0.3333$$

$$P(W = secret|C = HAM) = \frac{1}{15} \approx 0.0667$$

P(secret | *SPAM*) = $\theta_5$ P(secret | HAM) = $\theta_6$

*ss* number of 'secret' in Spam vs. *sh* number of 'secret' in Ham

*nss* number of non-'secret' in Spam vs. *nsh* number of non-'secret' in Ham

P(d| $h_{\theta,\theta5,\theta6}$ ) = $\theta^s \cdot (1-\theta)^h \cdot \theta_5^{ss} \cdot (1-\theta_5)^{ns-s} \cdot \theta_6^{sh} \cdot (1-\theta_6)^{ns-h}$

L = $[s \cdot \log\theta + h \cdot \log(1-\theta)] + [ss \cdot \log\theta_5 + (nss) \cdot \log(1-\theta_5)] + [sh \cdot \log\theta_6 + (nsh) \cdot \log(1-\theta_6)]$

$\frac{dL}{d\theta} = \frac{s}{\theta} - \frac{h}{1-\theta} = 0$ $\theta = \frac{s}{s+h} = \frac{3}{3+5} = \frac{3}{8} = 0.375$

$\frac{dL}{d\theta 5} = \frac{ss}{\theta 5} - \frac{nss}{1-\theta 5} = 0$ $\theta_5 = \frac{ss}{ss+nss} = \frac{3}{3+6} = \frac{3}{9} = \frac{1}{3} \approx$ .33 = P(secret | SPAM)

$\frac{dL}{d\theta 6} = \frac{sh}{\theta 6} - \frac{nsh}{1-\theta 6} = 0$ $\theta_6 = \frac{sh}{sh+nsh} = \frac{1}{1+14} = \frac{1}{15} \approx$ .0667 = P(secret | HAM)

(4) [10] Now, the new message "*golf*" is received. Compute the *likelihood* of this message is *SPAM*.

$$P(SPAM|W = golf)$$

$$= \frac{P(W = golf|SPAM) \times P(SPAM)}{P(W = golf|SPAM)P(SPAM) + P(W = golf|HAM)P(HAM)}$$

$$= \frac{1/9 \times 3/8}{1/9 \times 3/8 + 5/15 \times 5/8} = \frac{3}{18} \approx 0.1667$$

**The parameter values of $h_\theta$, $h_{\theta 11}$, $h_{,\theta 12}$ have to be learned by ML learning for their use above.**
**P(SPAM) = θ is learned in 1), θ = 0.375.**
**Similarly, learn $\theta_{11}$ = P(golf | SPAM) and $\theta_{12}$ = P(golf | HAM) using log-likelihood by ML learning.**

P(SPAM| golf, $h_{\theta,\theta 11,\theta 12}$)

= α P(SPAM| $h_{\theta,\theta 11,\theta 12}$)· P(golf | SPAM, $h_{\theta,\theta 11,\theta 12}$)

= α $\theta \cdot \theta_{11}$ Similarly, $\theta_{11}$ may be learned by ML learning, $= \frac{1}{9}$

$= \frac{3}{8} \cdot \frac{1}{9} = \frac{1}{24}\alpha$

P(HAM| golf, $h_{\theta,\theta 11,\theta 12}$)

= α P(HAM| $h_{\theta,\theta 11,\theta 12}$)· P(golf | HAM, $h_{\theta,\theta 11,\theta 12}$)

= α $(1-\theta) \cdot \theta_{12}$ Similarly, $\theta_{12}$ may be learned by ML learning, $= \frac{5}{15} = 1/3$

$= \frac{5}{8} \cdot \frac{1}{3} = \frac{5}{24}\alpha$

After normalizing the above, < 1/24, 5/24> = <.1667, .8333 >

(5) [10] The new message "*secret is secret*" is received. What is the *likelihood* of this message is *SPAM*?

P(C=SPAM|W=secret, is, secret)

$=\alpha\, P(W = \text{secret}|SPAM) \times P(W = \text{is}|SPAM) \times P(W = \text{secret}|SPAM) \times P(SPAM)$

$=\alpha \frac{1}{3} \times \frac{1}{9} \times \frac{1}{3} \times \frac{3}{8} = \frac{1}{216}\alpha$

P(C=HAM|W=secret, is, secret)

$=\alpha\, P(W = \text{secret}|HAM) \times P(W = \text{is}|HAM) \times P(W = \text{secret}|HAM) \times P(HAM)$

$=\alpha\frac{1}{15} \times \frac{1}{15} \times \frac{1}{15} \times \frac{5}{8} = \frac{1}{5400}\,\alpha$

So, $< \frac{1}{216}\alpha,\ \frac{1}{5400}\,\alpha > = < 5400\alpha, 216\alpha > =< 25\alpha, 1\alpha >$

$= < 25/26, 1/26 > = <.9615, .0285>$

i.e.

P(C=SPAM|W=secret, is, secret) =

$$\frac{P(W = \text{secret}|SPAM) \times P(W = \text{is}|SPAM) \times P(W = \text{secret}|SPAM) \times P(SPAM)}{P(\text{secret}|SPAM)P(\text{is}|SPAM)P(\text{secret}|SPAM)P(SPAM) + P(\text{secret}|HAM)P(\text{is}|HAM)P(secret|HAM)P(HAM)}$$

$$\approx \frac{\frac{1}{3}\times\frac{1}{9}\times\frac{1}{3}\times\frac{3}{8}}{\frac{1}{3}\times\frac{1}{9}\times\frac{1}{3}\times\frac{3}{8}+\frac{1}{15}\times\frac{1}{15}\times\frac{1}{15}\times\frac{5}{8}} = \frac{\frac{1}{216}}{\frac{1}{216}+\frac{1}{5400}} = \frac{5400}{216+5400} = \frac{5400}{5616} = \frac{25}{26} \approx 0.9615$$

(6) [10] For a new message, "*tomorrow is secret*", what is the likelihood of this message is *SPAM* and *HAM*, respectively?

P(C=SPAM|W=tomorrow, is, secret)

$= P(tomorrow|SPAM) \times P(\text{is}|SPAM) \times P(\text{secret}|SPAM) \times P(SPAM)\alpha$

$= 0 \times 1/9 \times 1/3 \times 3/8\alpha = 0\alpha = 0$

P(C=HAM|W=tomorrow, is, secret)

$= P(tomorrow|HAM) \times P(\text{is}|HAM) \times P(\text{secret}|HAM) \times P(HAM)\alpha$

$= \frac{2}{15} \times \frac{1}{15} \times \frac{1}{15} \times \frac{5}{8}\alpha = 1$

i.e.

P(C=SPAM|W=tomorrow, is, secret) =

$$\frac{P(tomorrow|SPAM) \times P(\text{is}|SPAM) \times P(\text{secret}|SPAM) \times P(SPAM)}{P(tomorrow|SPAM)P(\text{is}|SPAM)P(\text{secret}|SPAM)P(SPAM) + P(tomorrow|HAM)P(\text{is}|HAM)P(secret|HAM)P(HAM)}$$

$$\approx \frac{0 \times 1/9 \times 1/3 \times 3/8}{0 \times 1/9 \times 1/3 \times 3/8 + 2/15 \times 1/15 \times 1/15 \times 5/8} = 0$$

P(C=HAM|W=tomorrow, is, secret)

$$= \frac{P(tomorrow|HAM) \times P(\text{is}|HAM) \times P(\text{secret}|HAM) \times P(HAM)}{P(tomorrow|SPAM)P(\text{is}|SPAM)P(\text{secret}|SPAM)P(SPAM) + P(tomorrow|HAM)P(\text{is}|HAM)P(secret|HAM)P(HAM)}$$

$$= \frac{2/15 \times 1/15 \times 1/15 \times 5/8}{0 \times 1/9 \times 3/9 \times 3/8 + 2/15 \times 1/15 \times 1/15 \times 5/8} = 1$$

**Q4. [20] MAP Learning**

In the data of Candy Example of Chap. 20 (slide #6), 6 candies are unwrapped one by one and a flavor of each candy is as follows: d1 = lime, d2 = cherry, d3= cherry, d4 = lime, d5 = cherry, d6 = cherry.

(1) [10] By computing the posterior probability of each hypothesis, given the above six data, decide a ***Maximum A Posteriori hypothesis*** (**$h_{MAP}$**) where each $h_i$ is defined in the slide #6.
You have to show the essential computational steps if you compute them manually or using an excel sheet. Otherwise, write a program to compute them. – Refer to the handout of Candy Example on the blackboard and follow the submission instruction.

(a) Prior Probability of each hypothesis:

- $P(h_1) = .1$
- $P(h_2) = .2$
- $P(h_3) = .4$
- $P(h_4) = .2$
- $P(h_5) = .1$

| | Hypothesis Prior: $P(h_i)$ | Likelihood of data under each hypothesis | |
|---|---|---|---|
| h1 | .1 = 1/10 | P(d=lime \| h1) = 0 | P(d=cherry \| h1) = 1 |
| h2 | .2 = 1/5 | P(d=lime \| h2) =.25 = ¼ | P(d=cherry \| h2) =.75 = 3/4 |
| h3 | .4 = 2/5 | P(d=lime \| h3) = .5 = 1/2 | P(d=cherry \| h3) = .5 = 1/2 |
| h4 | .2 = 1/5 | P(d=lime \| h4) = .75 = 3/4 | P(d=cherry \| h4) = .25 = 1/4 |
| h5 | .1 = 1/10 | P(d=lime \| h5) = 1 | P(d=cherry \| h5) = 0 |

(b) The 1st candy is unwrapped: $d_1$ = lime

- $P(h_1|d_1=lime) = \alpha\, P(d_1= l|h_1)P(h_1) = \alpha\cdot 0\cdot 0.1 = 0 \quad \Rightarrow 0$
- $P(h_2|d_1=lime) = \alpha\, P(d_1= l\, |h_2)P(h_2) = \alpha\cdot 0.25\cdot 0.2 = 0.05\alpha \quad \Rightarrow 0.1$
- $P(h_3|d_1=lime) = \alpha\, P(d_1= l\, |h_3)P(h_3) = \alpha\cdot 0.5\cdot 0.4 = 0.20\alpha \quad \Rightarrow 0.4$
- $P(h_4|d_1=lime) = \alpha\, P(d_1= l\, |h_4)P(h_4) = \alpha\cdot 0.75\cdot 0.2 = 0.15\alpha \quad \Rightarrow 0.3$
- $P(h_5|d_1=lime) = \alpha\, P(d_1= l\, |h_5)P(h_5) = \alpha\cdot 1\cdot 0.1 = 0.1\alpha \quad \Rightarrow 0.2$

(c) The $2^{nd}$ candy is unwrapped: $d_2$ = cherry

- $P(h_1|d_1=l, d_2=c) = \alpha\, P(d_1= l\, , d_2= c\, |h_1)P(h_1) = \alpha\cdot 0\cdot 1\cdot 0.1 = 0 \quad \Rightarrow 0$
- $P(h_2|d_1=l, d_2=c) = \alpha\, P(d_1= l\, , d_2= c\, |h_2)P(h_2) = \alpha\cdot 0.25\cdot 0.75\cdot 0.2 = 0.0375\alpha \Rightarrow 0.2143$
- $P(h_3|d_1=l, d_2=c) = \alpha\, P(d_1= l\, , d_2= c\, |h_3)P(h_3) = \alpha\cdot 0.5\cdot 0.5\cdot 0.4 = 0.1\alpha \quad \Rightarrow 0.5714$
- $P(h_4|d_1=l, d_2=c) = \alpha\, P(d_1= l\, , d_2= c\, |h_4)P(h_4) = \alpha\cdot 0.75\cdot 0.25\cdot 0.2 = 0.0375\alpha \Rightarrow 0.2143$
- $P(h_5|d_1=l, d_2=c) = \alpha\, P(d_1= l\, , d_2= c\, |h_5)P(h_5) = \alpha\cdot 1\cdot 0\cdot 0.1 = 0\alpha \quad \Rightarrow 0$

(d) The $3^{rd}$ candy is unwrapped: $d_3$ = cherry

- $P(h_1|d_1=l, d_2=c, d_3=c) = \alpha\, P(d_1= l\, , d_2= c, d_3=c\, |h_1)P(h_1)$
  $= \alpha\cdot 0\cdot 1\cdot 1\cdot 0.1 = 0\alpha \quad \Rightarrow 0$
- $P(h_2|d_1=l, d_2=c, d_3=c) = \alpha\, P(d_1= l\, , d_2= c, d_3=c\, |h_2)P(h_2)$
  $= \alpha\cdot 0.25\, \cdot .75^2\cdot 0.2 = 0.028125\alpha \quad \Rightarrow 0.32143$
- $P(h_3|d_1=l, d_2=c, d_3=c) = \alpha\, P(d_1=l\, , d_2= c, d_3=c\, |h_3)P(h_3)$
  $= \alpha\cdot 0.5^3\cdot 0.4 = 0.05\alpha \quad \Rightarrow 0.57143$
- $P(h_4|d_1=l, d_2=c, d_3=c) = \alpha\, P(d_1= l\, , d_2= c, d_3=c\, |h_4)P(h_4)$
  $= \alpha\cdot 0.75\, \cdot 0.25^2\cdot 0.2 = 0.009375\alpha \quad \Rightarrow 0.10714$
- $P(h_5|d_1=l, d_2=c, d_3=c) = \alpha\, P(d_1= l\, , d_2= c, d_3=c\, |h_5)P(h_5)$
  $= \alpha\cdot 1\cdot 0\cdot 0\cdot 0.1 = 0\alpha \quad \Rightarrow 0$

(e) The $4^{th}$ is unwrapped: $d_4$ = lime.

- $P(h_1|d_1=l, d_2=c, d_3=c, d_4=l) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l |h_1)P(h_1)$
  $= \alpha \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 0.1 = 0\alpha \Rightarrow 0$
- $P(h_2|d_1=l, d_2=c, d_3=c, d_4=l) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l |h_2)P(h_2)$
  $= \alpha \cdot 0.25^2 \cdot .75^2 \cdot 0.2 = 0.00703\alpha \Rightarrow 0.18$
- $P(h_3|d_1=l, d_2=c, d_3=c, d_4=l) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l |h_3)P(h_3)$
  $= \alpha \cdot 0.5^4 \cdot 0.4 = 0.025\alpha \Rightarrow 0.64$
- $P(h_4|d_1=l, d_2=c, d_3=c, d_4=l) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l |h_4)P(h_4)$
  $= \alpha \cdot 0.75^2 \cdot 0.25^2 \cdot 0.2 = 0.00703\alpha \Rightarrow 0.18$
- $P(h_5|d_1=l, d_2=c, d_3=c, d_4=l) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l |h_5)P(h_5)$
  $= \alpha \cdot 1 \cdot 0 \cdot 0 \cdot 1 \cdot 0.1 = 0\alpha \Rightarrow 0$

(f) The $5^{th}$ is unwrapped: $d_5$ = cherry

- $P(h_1|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c |h_1)P(h_1)$
  $= \alpha \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 0.1 = 0\alpha \Rightarrow 0$
- $P(h_2|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c |h_2)P(h_2)$
  $= \alpha \cdot 0.25^2 \cdot .75^3 \cdot 0.2 = 0.00527\alpha \Rightarrow 0.27$
- $P(h_3|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c |h_3)P(h_3)$
  $= \alpha \cdot 0.5^5 \cdot 0.4 = 0.0125\alpha \Rightarrow 0.64$
- $P(h_4|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c |h_4)P(h_4)$
  $= \alpha \cdot 0.75^2 \cdot 0.25^3 \cdot 0.2 = 0.00176\alpha \Rightarrow 0.09$
- $P(h_5|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c |h_5)P(h_5)$
  $= \alpha \cdot 1 \cdot 0 \cdot 0 \cdot 1 \cdot 0 \cdot 0.1 = 0\alpha \Rightarrow 0$

(g) The $6^{th}$ is unwrapped: $d_6$ = cherry

- $P(h_1|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c |h_1)P(h_1)$
  $= \alpha \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0.1 = 0\alpha \Rightarrow 0$
- $P(h_2|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c |h_2)P(h_2)$
  $= \alpha \cdot 0.25^2 \cdot .75^4 \cdot 0.2 = 0.003956\alpha \Rightarrow 0.37156$
- $P(h_3|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c |h_3)P(h_3)$
  $= \alpha \cdot 0.5^6 \cdot 0.4 = 0.00625\alpha \Rightarrow 0.58716$
- $P(h_4|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c |h_4)P(h_4)$
  $= \alpha \cdot 0.75^2 \cdot 0.25^4 \cdot 0.2 = 0.000439\alpha \Rightarrow 0.04128$
- $P(h_5|d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c) = \alpha P(d_1=l, d_2=c, d_3=c, d_4=l, d_5=c, d_6=c |h_5)P(h_5)$
  $= \alpha \cdot 1 \cdot 0 \cdot 0 \cdot 1 \cdot 0 \cdot 0 \cdot 0.1 = 0\alpha \Rightarrow 0$

Since the posterior probability of h3, P(h3| d1, .. d6) is the highest among those of 5 hypotheses, $h_{MAP}$ = h3.

(2) [10] The $7^{th}$ candy is about to be unwrapped. Compute the prediction probability, <P(d7= cherry | $\boldsymbol{h_{MAP}}$), P(d7= lime | $\boldsymbol{h_{MAP}}$) > based on $\boldsymbol{h_{MAP}}$ in (1), and give a prediction on the flavor of the $7^{th}$ candy.

P(d7= cherry | $h_{MAP}$) = P($d_7$=cherry | $\mathbf{h_3}$) = 0.5 and
P(d7= lime | $h_{MAP}$) = P($d_7$=lime | $\mathbf{h_3}$) = 0.5 whose likelihood is given in (a).

<P(d7= cherry | $\boldsymbol{h_{MAP}}$), P(d7= lime | $\boldsymbol{h_{MAP}}$) > = < 0.5, 0.5 >