CS 229 - Machine Learning
Xiangliang Zhang
Computer Science (CS)/Statistics (STAT) Program
Computer, Electrical and Mathematical Sciences & Engineering (CEMSE) Division
King Abdullah University of Science and Technology (KAUST)

HOMEWORK IV

Henrique Aparecido Laureano

Spring Semester 2018

Contents

Question 1: Fisher's criterion (Exercise 4.5 of Bishop's book)

By making use of (4.20), (4.23), and (4.24), show that the Fisher criterion (4.25) can be written in the form (4.26).

Solution:

$$\begin{aligned}\text{Fisher criterion :} \quad J(\mathbf{w}) &= \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2} \qquad (4.25)\\ &= \frac{(\mathbf{w}^\top \mathbf{m}_2 - \mathbf{w}^\top \mathbf{m}_1)^2}{\sum_{n \in C_1} (\mathbf{w}^\top \mathbf{x}_n - \mathbf{w}^\top \mathbf{m}_1)^2 + \sum_{n \in C_2} (\mathbf{w}^\top \mathbf{x}_n - \mathbf{w}^\top \mathbf{m}_2)^2},\end{aligned}$$

applying

$$\begin{aligned} m_k &= \mathbf{w}^\top \mathbf{m}_k, && (4.23)\\ s_k^2 &= \sum_{n \in C_k} (y_n - m_k)^2 \quad \text{and} && (4.24)\\ y &= \mathbf{w}^\top \mathbf{x}. && (4.20)\end{aligned}$$

Returning to J,

$$\begin{aligned} J(\mathbf{w}) &= \frac{(\mathbf{w}^\top \mathbf{m}_2 - \mathbf{w}^\top \mathbf{m}_1)^2}{\sum_{n \in C_1} (\mathbf{w}^\top \mathbf{x}_n - \mathbf{w}^\top \mathbf{m}_1)^2 + \sum_{n \in C_2} (\mathbf{w}^\top \mathbf{x}_n - \mathbf{w}^\top \mathbf{m}_2)^2}\\ &= \frac{\mathbf{w}^\top (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^\top \mathbf{w}}{\mathbf{w}^\top \Big[\sum_{n \in C_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^\top + \sum_{n \in C_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^\top \Big] \mathbf{w}}\\ &= \frac{\mathbf{w}^\top \mathbf{S}_\mathrm{B} \mathbf{w}}{\mathbf{w}^\top \mathbf{S}_\mathrm{W} \mathbf{w}}, \qquad (4.26)\end{aligned}$$

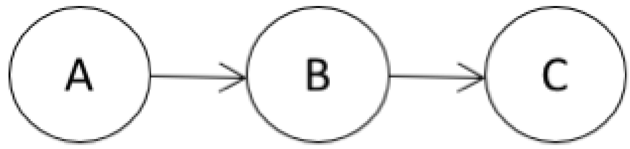
with

$$\begin{aligned} \mathbf{S}_\mathrm{B} &= (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^\top \quad \text{and}\\ \mathbf{S}_\mathrm{W} &= \sum_{n \in C_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^\top + \sum_{n \in C_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^\top .\end{aligned}$$

□

Question 2: Bayes Net - Proofs

Given this Bayes net



Prove that C is conditionally independent of A, given B. You should use the definition of joint probabilities for a Bayes net.

Solution:

$$p(C, A \mid B) = \frac{p(C, A, B)}{p(B)} = \frac{p(A)p(B \mid A)p(C \mid B)}{p(B)} = \frac{p(A, B)p(C \mid B)}{p(B)} = p(A \mid B)p(C \mid B)$$
$$\Rightarrow C \perp\!\!\!\perp A \mid B.$$

□

Question 3: Bayesian Network

We are going to take the perspective of an instructor who wants to determine whether a student has understood the material, based on the exam score. Figure 2 gives a Bayes net for this. As you can see, whether the student scores high on the exam is influenced both by whether she is a good test taker, and whether she understood the material. Both of those, in turn, are influenced by whether she is intelligent; whether she understood the material is also influenced by whether she is a hard worker.

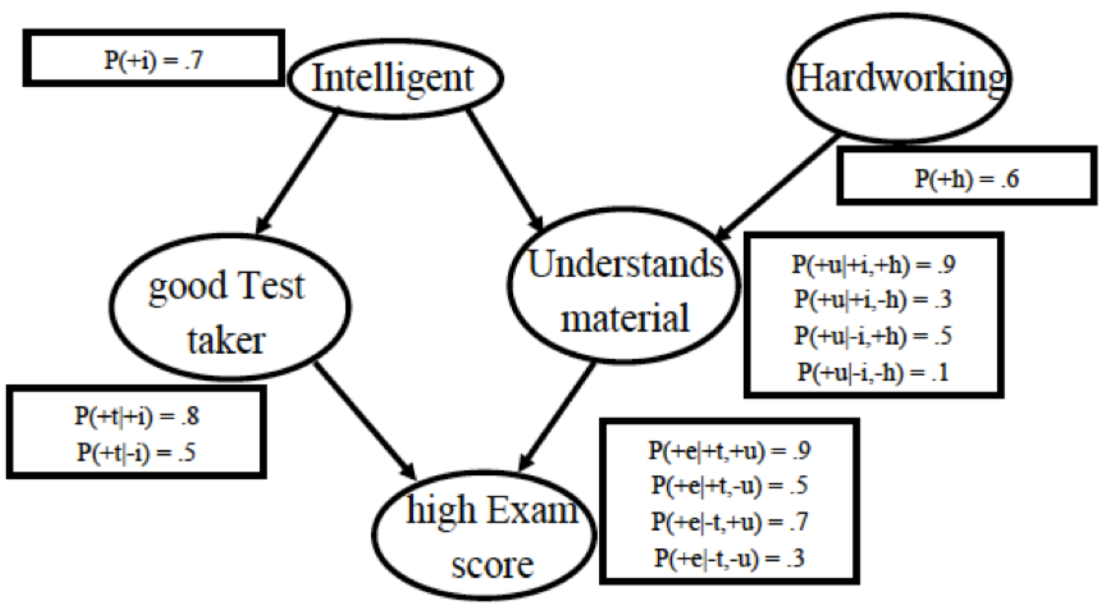


Figure 2: A Bayesian network representing what influences an exam score.

(1)

Using variable elimination (by hand!), compute the probability that a student who did well on the test actually understood the material, that is, compute P(+ u | + e).

Solution:

Renaming the states for convenience:

- Intelligent: I,
- Hardworking : H,
- good Test taker: T,
- Understands material: U,
- high Exam score: E.

We are interested in

$$\mathbb{P}(\mathrm{U} \mid \mathrm{E}) = \frac{\mathbb{P}(\mathrm{U}, \mathrm{E})}{\mathbb{P}(\mathrm{E})}$$

and we have 6 initial factors (giving by the Bayesian network representation in Figure 2):

$$\mathbb{P}(\mathrm{I}), \quad \mathbb{P}(\mathrm{H}), \quad \mathbb{P}(\mathrm{T} \mid \mathrm{I}), \quad \mathbb{P}(\mathrm{U} \mid \mathrm{I}, \mathrm{H}), \quad \mathbb{P}(\mathrm{E} \mid \mathrm{T}, \mathrm{U}).$$

To compute first $\mathbb{P}(\mathrm{U}, \mathrm{E})$, following the elimination order I, H, T, we have

$$\begin{aligned}
f_1(\mathrm{H}, \mathrm{T}, \mathrm{U}) &= \sum_{\mathrm{i}} \mathbb{P}(\mathrm{i})\mathbb{P}(\mathrm{T} \mid \mathrm{i})\mathbb{P}(\mathrm{U} \mid \mathrm{i}, \mathrm{H}) = 0.672 \\
f_2(\mathrm{T}, \mathrm{U}) &= \sum_{\mathrm{h}} \mathbb{P}(\mathrm{h}) f_1(\mathrm{h}, \mathrm{T}, \mathrm{U}) = 0.4032 \\
f_3(\mathrm{E}, \mathrm{U}) &= \sum_{\mathrm{t}} \mathbb{P}(\mathrm{E} \mid \mathrm{t}, \mathrm{U}) f_2(\mathrm{t}, \mathrm{U}) = 0.56448 \quad \Rightarrow \mathbb{P}(\mathrm{U}, \mathrm{E}).
\end{aligned}$$

Now, $\mathbb{P}(\mathrm{E})$:

$$\begin{aligned}
\mathbb{P}(\mathrm{I}, \mathrm{H}, \mathrm{T}, \mathrm{U}, \mathrm{E}) &= \mathbb{P}(\mathrm{I})\mathbb{P}(\mathrm{T} \mid \mathrm{I})\mathbb{P}(\mathrm{H})\mathbb{P}(\mathrm{U} \mid \mathrm{I}, \mathrm{H})\mathbb{P}(\mathrm{E} \mid \mathrm{T}, \mathrm{U}) \\
\mathbb{P}(\mathrm{E}) &= \sum_{\mathrm{i}} \sum_{\mathrm{h}} \sum_{\mathrm{t}} \sum_{\mathrm{u}} \mathbb{P}(\mathrm{I})\mathbb{P}(\mathrm{T} \mid \mathrm{I})\mathbb{P}(\mathrm{H})\mathbb{P}(\mathrm{U} \mid \mathrm{I}, \mathrm{H})\mathbb{P}(\mathrm{E} \mid \mathrm{T}, \mathrm{U}) \\
f_{\mathrm{h}}(\mathrm{U} \mid \mathrm{I}) &= \sum_{\mathrm{h}} \mathbb{P}(\mathrm{h})\mathbb{P}(\mathrm{U} \mid \mathrm{I}, \mathrm{h}) \\
\mathbb{P}(\mathrm{E}) &= \sum_{\mathrm{i}} \sum_{\mathrm{t}} \sum_{\mathrm{u}} f_{\mathrm{h}}(\mathrm{U} \mid \mathrm{I})\mathbb{P}(\mathrm{I})\mathbb{P}(\mathrm{T} \mid \mathrm{I})\mathbb{P}(\mathrm{E} \mid \mathrm{T}, \mathrm{U}) \\
f_{\mathrm{i}}(\mathrm{U}, \mathrm{T}) &= \sum_{\mathrm{i}} \mathbb{P}(\mathrm{i}) f_{\mathrm{h}}(\mathrm{U} \mid \mathrm{i})\mathbb{P}(\mathrm{T} \mid \mathrm{i}) \\
\mathbb{P}(\mathrm{E}) &= \sum_{\mathrm{t}} \sum_{\mathrm{u}} f_{\mathrm{i}}(\mathrm{u}, \mathrm{t})\mathbb{P}(\mathrm{E} \mid \mathrm{t}, \mathrm{u}) \\
&= f_{\mathrm{t},\,\mathrm{u}}(\mathrm{E}) \\
&= 0.9.
\end{aligned}$$

Therefore,

$$\mathbb{P}(\mathrm{U} \mid \mathrm{E}) = \frac{\mathbb{P}(\mathrm{U}, \mathrm{E})}{\mathbb{P}(\mathrm{E})} = \frac{0.56448}{0.9} = 0.6272.$$

The probability that a student who did well on the test actually understood the material is 0.6272.

□

(2)

For the above Bayesian network, label the following statements about conditional independence as true or false. For this question, you should consider only the structure of the Bayesian network, not the specific probabilities. Explain each of your answers.

1) T and U are independent.

False.

$$\begin{aligned}
p(t,u,i,h) &= p(i)p(h)p(t \mid i)p(u \mid i,h) \\
p(t,u) &= \sum_i \sum_h p(t \mid i)p(i)p(u \mid i,h)p(h) \\
p(t,u) &= \sum_i p(t \mid i)p(i) \sum_h p(u \mid i,h)p(h) \\
p(t,u) &= \sum_i p(t \mid i)p(i)p(u \mid i) \\
& t \not\perp\!\!\!\perp u \mid \varnothing
\end{aligned}$$

2) T and U are conditionally independent given I, E, and H.

False.

$$\begin{aligned}
p(t,u \mid i,e,h) &= \frac{p(t,u,i,e,h)}{p(i,e,h)} \\
&= \frac{p(i)p(h)p(t \mid i)p(u \mid i,h)p(e \mid t,u)}{p(i)p(h)p(e \mid t,u)} \\
&= p(t \mid i)p(u \mid i,h) \\
& t \perp\!\!\!\perp u \mid i,h
\end{aligned}$$

T and U are conditionally independent given I and H, not given I, E, and H.

3) T and U are conditionally independent given I and H.

True.

$$
\begin{aligned}
p(t, u \mid i, h) &= \frac{p(t, u, i, h)}{p(i, h)} \\
&= \frac{p(i)p(h)p(t \mid i)p(u \mid i, h)}{p(i)p(h)} \\
&= p(t \mid i)p(u \mid i, h) \\
& t \perp\!\!\!\perp u \mid i, h
\end{aligned}
$$

4) E and H are conditionally independent given U.

True.

$$
\begin{aligned}
p(e, h \mid u) &= \frac{p(e, h, u)}{p(u)} \\
&= \frac{p(h)p(u \mid i, h)p(e \mid t, u)}{p(u)}
\end{aligned}
$$

Sorry, no enough time to complete the homework, I have weekly homeworks in the four courses that I'm doing and this week I was not able to finish this.

5) E and H are conditionally independent given U, I, and T.

6) I and H are conditionally independent given E.

7) I and H are conditionally independent given T.

8) T and H are independent.

9) T and H are conditionally independent given E.

10) T and H are conditionally independent given E and U.

□

Question 4: k-nearest neighbor

Write your k-nn code to classify

(1)

Students in data set in `test_data_binary.txt` by training data in `logreg_data_binary.txt`.

(2)

Students in data set in `test_data_3class.txt` **by training data in** `logreg_data_3class.txt`**.**

Report how many student labels in test data are correctly predicted.

NOTE: choose an appropriate k to reach the best prediction.

■
