Machine Learning 2 - Homework 7

Pascal M. Esser

May 21, 2018

Collaborators: Sindy Löwe

1

1.1

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{old}) = -\frac{N-1}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\sum_{n=2}^{N}(\boldsymbol{z}_n - \boldsymbol{A}\boldsymbol{z}_{n-1})^\top\boldsymbol{\Gamma}^{-1}(\boldsymbol{z}_n - \boldsymbol{A}\boldsymbol{z}_{n-1})\right] + C \tag{1}$$

$$= -\frac{N-1}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\sum_{n=2}^{N}\boldsymbol{z}_n^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{z}_n - 2\boldsymbol{z}_n^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\boldsymbol{z}_{n-1} + \boldsymbol{z}_{n-1}^\top\boldsymbol{A}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\boldsymbol{z}_{n-1}\right] + C \tag{2}$$

$$\frac{\partial\mathcal{Q}}{\partial\boldsymbol{A}} = \text{[using: push } \mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}[\cdot] \text{ in by linearity of expectation, Matrix cookbook 88]} \tag{3}$$

$$= -\frac{1}{2}\sum_{n=2}^{N} -2\boldsymbol{\Gamma}^{-1}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_n^\top\boldsymbol{z}_{n-1}\right] + 2\boldsymbol{\Gamma}^{-1}\boldsymbol{A}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_{n-1}\boldsymbol{z}_{n-1}^\top\right] = 0 \tag{4}$$

$$\text{[using: } -2\boldsymbol{\Gamma}^{-1} \text{ and } \frac{1}{2} \text{ cancels out]} \tag{5}$$

$$\boldsymbol{A}^{new} = \left(\sum_{n=2}^{N}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_n^\top\boldsymbol{z}_{n-1}\right]\right)\left(\sum_{n=2}^{N}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_{n-1}\boldsymbol{z}_{n-1}^\top\right]\right)^{-1} \tag{6}$$

$$\frac{\partial\mathcal{Q}}{\partial\boldsymbol{\Gamma}} = \text{[using: derive from Equation 1, Matrix cookbook 57 and 61]} \tag{7}$$

$$= -\frac{N-1}{2}\boldsymbol{\Gamma}^{-1} - \frac{1}{2}\left[-\boldsymbol{\Gamma}^{-1}\sum_{n=2}^{N}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[(\boldsymbol{z}_n - \boldsymbol{A}\boldsymbol{z}_{n-1})(\boldsymbol{z}_n - \boldsymbol{A}\boldsymbol{z}_{n-1})^\top\right]\boldsymbol{\Gamma}^{-1}\right] \tag{8}$$

$$\text{[using: multiply } \boldsymbol{\Gamma} \text{ on the right]} \tag{9}$$

$$\boldsymbol{\Gamma}^{new} = \frac{1}{N-1}\sum_{n=2}^{N}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[(\boldsymbol{z}_n - \boldsymbol{A}\boldsymbol{z}_{n-1})(\boldsymbol{z}_n - \boldsymbol{A}\boldsymbol{z}_{n-1})^\top\right] \tag{10}$$

$$\text{which we can also rewrtie again in terms of } \boldsymbol{A}^{new} \tag{11}$$

$$= \frac{1}{N-1}\sum_{n=2}^{N}\left[\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_n\boldsymbol{z}_n^\top\right] - \boldsymbol{A}^{new}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_{n-1}\boldsymbol{z}_n^\top\right]\right. \tag{12}$$

$$\left. - \mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_n\boldsymbol{z}_{n-1}^\top\right]\boldsymbol{A}^{new\top} - \boldsymbol{A}^{new}\mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_{n-1}\boldsymbol{z}_{n-1}^\top\right]\boldsymbol{A}^{new\top}\right] \tag{13}$$

1.2

We can directly reuse the results from § 1.1 as both models based on Gaussians just with different parameterizations.

$$\boldsymbol{C}^{new} = \left(\sum_{n=1}^{N} \boldsymbol{x}_n^\top \mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_n\right]\right)\left(\sum_{n=1}^{N} \mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[\boldsymbol{z}_n^\top \boldsymbol{z}_n\right]\right)^{-1} \tag{14}$$

$$\boldsymbol{\Sigma}^{new} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{E}_{\boldsymbol{z}|\boldsymbol{\theta}^{old}}\left[(\boldsymbol{z}_n - \boldsymbol{C}\boldsymbol{z}_{n-1})(_n - \boldsymbol{C}\boldsymbol{z}_n)^\top\right] \tag{15}$$

2

a) See Figure 1

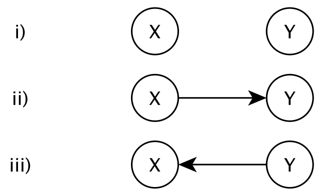

Figure 1: all different possible structures

b)

$$i)\ p(X,Y) = p(X)p(Y) \tag{16}$$
$$ii)\ p(X,Y) = p(X)p(Y|X) \tag{17}$$
$$iii)\ p(X,Y) = p(X|Y)p(Y) \tag{18}$$
$$\tag{19}$$

c)

$$i)\ p(Y|X) = p(Y) = \frac{p(X,Y)}{p(X)} \tag{20}$$
$$ii)\ p(Y|X) = \frac{p(X,Y)}{p(X)} \tag{21}$$
$$iii)\ p(Y|X) = \frac{p(X|Y)p(Y)}{\int p(X|Y)p(Y)dY} \tag{22}$$
$$\tag{23}$$

$$i)\ p(Y|do(X)) = p(Y) \tag{24}$$
$$ii)\ p(Y|do(X)) = \frac{p(X,Y)}{p(X)} \tag{25}$$
$$iii)\ p(Y|do(X)) = p(Y) \tag{26}$$
$$\tag{27}$$

e) $p(Y|X) = p(cancer|smokes)$ probability that someone has cancer given we observe smoking
$p(Y|X) = p(cancer|do(smokes))$ probability that someone has cancer given that we force the person to smoke

3

For this task let D = Drugs, R = Recovery

1. a) $p(R|D=1) = 20/40 = 50\%, p(R|D=0) = 16/40 = 40\%$

   b) Yes as the recovery rate for patients taking the drug is higher

2. a) Males $p(R|D=1) = 18/30 = 60\%, p(R|D=0) = 7/10 = 70\%$
   Females $p(R|D=1) = 2/10 = 20\%, p(R|D=0) = 9/30 = 30\%$

   b) For both cases: no as the recovery rate in the control group is higher

3. without knowing the gender we recommend the drug but after knowing the gender we would advice patient of both genders to not take the drug. Therefore the results seem to be not conclusive and we could advise against taking the drug as thy always bring side effects and we cant conclusively make a statement about if it is helping.

4. a) $S = \{M\}$ blocks the only back-door path therefore: $p(R|\mathbf{do}(D)) = \int p(R|D,M)p(M)dM$

   b) we can write $p(R|D) = \int p(R|D,M)p(M|D)dM$ and therefore $p(R|\mathbf{do}(D)) \neq p(R|D)$

   c)

$$\begin{aligned}
p(R=1|\mathbf{do}(D=1)) &= \sum_M p(R=1|D=1,M)p(M) && (28)\\
&= 0.6 * 0.5 + 0.2 * 0.5 && (29)\\
&= 0.4 && (30)\\
&\leq p(R=1|\mathbf{do}(D=0)) && (31)\\
&= \sum_M p(R=1|D=0,M)p(M) && (32)\\
&= 0.7 * 0.5 + 0.3 * 0.5 && (33)\\
&= 0.5 && (34)
\end{aligned}$$

   therefor don't take the drug (we could also keep $p(M)$ in the formula above without writing it out but would come to the same results as it cancels out.)

5. a) back-door path is the empty set and therefore $p(R|\mathbf{do}(D)) = \int p(R|D,M)p(M|D)dM = p(R|D)$

   b) using the results from 4) and 5a): yes

   c) yes as $p(R=1|\mathbf{do}(D=1)) = p(R=1|D=1) = 0.5 > 0.4 = p(R=1|D=0) = p(R=1|\mathbf{do}(D=0))$

6. a) L_1 :How responsible is the person - if he is responsible he will take the drugs on a regular basis. Not observable.
   L_2, How good is the doctor. It does not influence if the patient takes the drugs or not but it influences the recovery for example due to the quality of the diagnosis or the quality of a surgery
   M : Time spend in the hospital. Influence by how responsible the person is as he will come in if he has to but also by how good your doctor is. If he is bad he might let the patient come in for unnecessary tests and the patient spends more time in the hospitable.

   b) back-door path is the empty set and therefore $p(R|\mathbf{do}(D)) = p(R|D)$

c) yes

d) take the drugs as gender is not part of the model and therefore the results from 5 apply: yes

4

1. $p(R=1)=p(E_R)=0.7, p(S=1)=p(E_S)=0.4, p(W|R,S)=R\vee S=\frac{p(R,S,W)}{p(R)p(S)}, p(R,S,W)=p(R)p(S)p(W|R,S), p(W=0)=p(S=0)p(R=0)=0.3*0.6=0.18$

2.
$$p(R=1|W=1)=\frac{p(R=1,W=1)}{p(W=1)}=\frac{\sum_S p(R=1,S,W=1)}{\sum_{S,R} p(R,S,W=1)} \tag{35}$$
$$=\frac{0.42+0.28}{0.12+0.42+0.28}=\frac{0.7}{0.82}=0.854 \tag{36}$$

3. no. causality does not equal correlation

4. remove all incoming edges which gives us Figure 2 with $E_R \perp\!\!\!\perp E_S$:

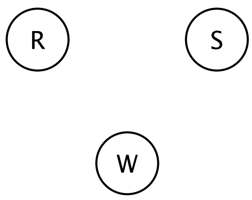

Figure 2

and therefore:

$$R=E_R \tag{37}$$
$$S=E_S \tag{38}$$
$$W=1 \tag{39}$$

5. $p(R|do(W=w))=p(R)=0.7$ and therefor in general $p(R|do(W=w))\neq p(R|W=w)$ .in specific here: $(p(R=1|W=1)=0.854, p(R=1|W=0)=0.146)$

6. the new causal model is given by Figure 3

$$R=E_R \tag{40}$$
$$S=s \tag{41}$$
$$W=R\vee S \tag{42}$$

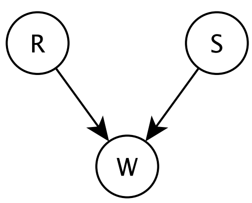

Figure 3

7. Since $(W \perp\!\!\!\perp S)_{\mathcal{G}_{\underline{S}}}$ holds we get $p(W|do(S = s)) = p(W|S = s)$ and also in the general case $p(W|do(S = s)) \neq p(W)$ we can use the action/observation exchange rule from do-calculus to get here in specific: $p(W = 1|S = 1) = 1, p(W = 1|S = 0) = 0.7, p(W = 0|S = 1) = 0, p(W = 0|S = 0) = 0.3, p(W = 1) = 0.82, p(W = 0) = 0.18$

5

1.

$$p(Y|\mathbf{do}(X), \boldsymbol{X}_{pa(X)}) = \frac{p(Y, \boldsymbol{X}_{pa(X)}|\mathbf{do}(X))}{p(\boldsymbol{X}_{pa(X)}|\mathbf{do}(X))} \tag{43}$$

$$\text{using truncated factorization theorem} \tag{44}$$

$$= \frac{\frac{p(Y, X, \boldsymbol{X}_{pa(X)})}{p(X|\boldsymbol{X}_{pa(X)})}}{\frac{p(X, \boldsymbol{X}_{pa(X)})}{p(X|\boldsymbol{X}_{pa(X)})}} \tag{45}$$

$$= \frac{p(Y, X, \boldsymbol{X}_{pa(X)})}{p(X, \boldsymbol{X}_{pa(X)})} \tag{46}$$

$$= p(Y|X, \boldsymbol{X}_{pa(X)}) \tag{47}$$

2.

$$p(Y|\mathbf{do}(X)) = \int p(Y|\mathbf{do}(X), \boldsymbol{X}_{pa(X)}) p(\boldsymbol{X}_{pa(X)}|\mathbf{do}(X)) d\boldsymbol{X}_{pa(X)} \tag{48}$$

$$\text{with } p(\boldsymbol{X}_{pa(X)}|\mathbf{do}(X)) = p(\boldsymbol{X}_{pa(X)}) \tag{49}$$

$$= \int p(Y|X, \boldsymbol{X}_{pa(X)}) p(\boldsymbol{X}_{pa(X)}) d\boldsymbol{X}_{pa(X)} \tag{50}$$