Solutions to problems of Chapter 4

Vahan Arsenyan

July 23, 2023

Problem 4.15 (ANMs): a) Consider the SCM

$$X := N_X$$
$$Y := 2X + N_Y$$

with N_X uniformly distributed between 1 and 3 and N_Y uniformly distributed between -0.5 and 0.5 and independent of N_X. The distribution $P_{X,Y}$ admits an ANM from X to Y. Draw the support of the joint distribution of X, Y and convince yourself that $P_{X,Y}$ does not admit an ANM from Y to X, that is there is no function g and independent noise variables M_X and M_Y such that

$$X = g(Y) + M_X$$
$$Y = M_Y$$

with M_X independent of M_Y.
b) Similarly as in part a), consider the SCM

$$X := N_X$$
$$Y := X^2 + N_Y$$

with N_X uniformly distributed between 1 and 3 and N_Y uniformly distributed between -0.5 and 0.5 and independent of N_X. Again, draw the support of $P_{X,Y}$ and convince yourself that there is no ANM from Y to X.
Solution: Please see the corresponding Jupyter notebook.
Problem 4.16 (Maximum likelihood): Assume that we are given an i.i.d. data set $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the model $Y = f(X) + N_Y$, with $X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right)$, and $N_Y \sim \mathcal{N}\left(\mu_{N_Y}, \sigma_{N_Y}^2\right)$ independent, where the function f is supposed to be known.
a) Prove that $f(x) = \mathbb{E}[Y \mid X = x]$.
b) Write $\mathbf{x} := (x_1, \ldots, x_n), \mathbf{y} := (y_1, \ldots, y_n)$ and consider the log-likelihood function

$$\ell_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = \ell_{\boldsymbol{\theta}}\left((x_1, y_1), \ldots, (x_n, y_n)\right) = \sum_{i=1}^{n} \log p_\theta\left(x_i, y_i\right),$$

where p_θ is the joint density over (X, Y) and $\theta := \left(\mu_X, \mu_{N_Y}, \sigma_X^2, \sigma_{N_Y}^2\right)$. Prove that for some $c_1, c_2 \in \mathbb{R}$ with $c_2 > 0$

$$\max_{\boldsymbol{\theta}} \ell_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = c_2 \cdot \left(c_1 - \log \widehat{\operatorname{var}}[\mathbf{x}] - \log \widehat{\operatorname{var}}[\mathbf{y} - f(\mathbf{x})]\right),$$

where $\widehat{\operatorname{var}}[\mathbf{z}] := \frac{1}{n}\sum_{i=1}^n \left(z_i - \frac{1}{n}\sum_{k=1}^n z_k\right)^2$ estimates the variance. Equation (4.20) motivates the comparison of expressions (4.18) and (4.19). The main difference is that in this exercise, we have used the conditional mean and not the outcome of the regression method. One can show that, asymptotically, the latter still produces correct results [Bühlmann et al., 2014].

Solution:

a) $\mathbb{E}[Y \mid X = x] = \mathbb{E}[f(X) + N_Y \mid X = x] = f(x) + \mu_{N_Y}$

b) First note that $p(x, y) = p(x)p(y \mid x)$ hence $\log p(x, y) = \log p(x) + \log p(y \mid x)$. Both $p(x)$ and $p(y|x)$ are Gaussian distributions. If one writes down the log likelihood for these distributions, they will get:

$$\log p(x) = -\log\sqrt{2\pi} - \log\sigma_x - \frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x}\right)^2$$
$$\log p(y \mid x) = -\log\sqrt{2\pi} - \log\sigma_{N_Y} - \frac{1}{2}\left(\frac{y - f(x) - \mu_{N_Y}}{\sigma_{N_Y}}\right)^2$$

plugging those equations into the log-likelihood function we get:

$$\sum_{i=1}^n \log p_\theta(x_i, y_i) = -2n\log\sqrt{2\pi} - n\log\sigma_X - n\log\sigma_{N_Y}$$
$$-\frac{1}{2}\sum_{i=1}^n\left[\left(\frac{x_i - \mu_x}{\sigma_X}\right)^2 + \left(\frac{y_i - f(x_i) - \mu_{N_Y}}{\sigma_{N_Y}}\right)^2\right] = n(-2\log\sqrt{2\pi} - \log\sigma_X - \log\sigma_{N_Y}$$
$$-\frac{1}{2}\left[\frac{1}{\sigma_X^2 n}\sum_{i=1}^n (x_i - \mu_X)^2 + \frac{1}{\sigma_{N_Y}^2 n}\sum_{i=1}^n (y_i - f(x_i) - \mu_{N_Y})^2\right])$$

replacing $\sigma_X, \sigma_{N_Y}, \mu_X, \mu_{N_Y}$ with their maximum likelihood estimates (denoted by hat), one may notice that summands in the square brackets become equal to 1. Truly:

$$\frac{1}{\hat\sigma_X^2 n}\sum_{i=1}^n (x_i - \hat\mu_X)^2 = \frac{1}{\hat\sigma_X^2}\cdot\frac{\sum_{i=1}^n (x_i - \hat\mu_X)^2}{n} = \frac{\hat\sigma_X^2}{\hat\sigma_X^2} = 1$$

and

$$\frac{1}{\hat\sigma_{N_Y}^2 n}\sum_{i=1}^n (y_i - f(x_i) - \hat\mu_{N_Y})^2 = \frac{1}{\hat\sigma_{N_Y}^2}\cdot\frac{\sum_{i=1}^n (y_i - f(x_i) - \hat\mu_{N_Y})^2}{n} = \frac{\hat\sigma_{N_Y}^2}{\hat\sigma_{N_Y}^2} = 1$$

Hence,

$$\max_{\boldsymbol{\theta}} \ell_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = n\left(-2\log\sqrt{2\pi} - \log\hat\sigma_X - \log\hat\sigma_{N_Y} - 1\right) = c_2(c_1 - \log\hat\sigma_X - \log\hat\sigma_{N_Y})$$