CDS: Machine Learning

Graphical Models extra excercises

a.)

$$p(x_i|y) = \frac{p(x_i, y)}{p(y)}$$

$$= \frac{\sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} p(x_1,\dots,x_n, y)}{\sum_{x_1,\dots,x_n} p(x_1,\dots,x_n, y)}$$

$$= \frac{\sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n} p(x_1)\dots p(x_n)p(y|x_1,\dots,x_n)}{\sum_{x_1,\dots,x_n} p(x_1)\dots p(x_n)p(y|x_1,\dots,x_n)}$$

notice: $\sum_{x_i,x_j}$ means $\sum_{x_i}\sum_{x_j}$, so above we have in the nominator $n-1$ summations and n in the denominator.

b.1

$$p(x_i \mid y, x_1 = 1) = \frac{p(x_i, y \mid x_1 = 1)}{p(y \mid x_1 = 1)}$$

$$= \frac{\sum_{x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} p(x_2, \ldots, x_n, y \mid x_1 = 1)}{\sum_{x_2, \ldots, x_n} p(x_2, \ldots, x_n, y \mid x_1 = 1)}$$

$$= \frac{\sum_{x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} p(x_2) \ldots p(x_n)\, p(y \mid x_1 = 1, x_2, \ldots, x_n)}{\sum_{x_2, \ldots, x_n} p(x_2) \ldots p(x_n)\, p(y \mid x_1 = 1, x_2, \ldots, x_n)}$$

Now gaussians :

a.) $p(x_1, x_2, y) = \mathcal{N}(x_1|0,1)\,\mathcal{N}(x_2|0,1)\,\mathcal{N}(y|\gamma(x_1+x_2), \sigma^2)$

$= \mathcal{N}(\vec{x}\,|\,\vec{\mu}, \Sigma)$

where :

$\vec{x} = (x_1, x_2, y)$. The only thing left to do

is to specify $\vec{\mu}$ and Σ.

we know $\mu_1 = \mu_{x_1} = 0$

$\mu_2 = \mu_{x_2} = 0$

$\mu_3 = \mu_y$: (for convenience : ksi $= \epsilon$)

$\mathbb{E}[y] = \mathbb{E}[\gamma(x_1+x_2) + \epsilon]$

$= \gamma\,\mathbb{E}[x_1] + \gamma\,\mathbb{E}[x_2] + \mathbb{E}[\epsilon]$

$= 0$

$\rightarrow \vec{\mu} = (0,0,0)$; $\Sigma_{11} = \Sigma_{22} = \mathbb{E}[x_2 x_2] = 1$

$\Sigma_{ij} \equiv \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$

$\Sigma_{12} = \Sigma_{21} = \mathbb{E}[x_1 x_2] = 0$ (they are assumed to be none-correlated within the question.)

$\Sigma_{33} = \mathbb{E}[y^2]$

$= \gamma^2\,\mathbb{E}[x_1^2 + x_2^2 + 2x_1x_2] + 2\gamma\,\mathbb{E}[\epsilon x_1 + \epsilon x_2]$

$+ \mathbb{E}[\epsilon^2] = 2\gamma^2 + \sigma^2$

$\Sigma_{12} = \Sigma_{21} = \Sigma_{31} = \Sigma_{13}$:

$\Sigma_{13} = \mathbb{E}[x_1(\gamma(x_1+x_2) + \epsilon)]$

$= \gamma\,\mathbb{E}[x_1^2] + \gamma\,\mathbb{E}[x_2 x_1] + \mathbb{E}[\cancel{x_1}\epsilon]$

$= \gamma$

$$\rightarrow \Sigma = \begin{pmatrix} 1 & 0 & \gamma \\ 0 & 1 & \gamma \\ \gamma & \gamma & 2\gamma^2+\sigma^2 \end{pmatrix}$$

b.)

we now use the definition of conditional gaussians

with:

$$\Sigma_{aa} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \cancel{\Sigma_{ab} = (\gamma, \gamma)} \qquad \Sigma_{ab} = \begin{pmatrix} \gamma \\ \gamma \end{pmatrix}$$

$$\Sigma_{bb} = 2\gamma^2 + \sigma^2 \qquad \cancel{\Sigma_{ba} = \begin{pmatrix} \gamma \\ \gamma \end{pmatrix}} \qquad \Sigma_{ba} = (\gamma, \gamma)$$

$$\mu_a = (0,0)^T \qquad \mu_b = 0$$

$$\rightarrow \quad p(x_1, x_2 | y) = \mathcal{N}((x_1, x_2) \mid \vec{\mu}_{a|b}, \Sigma_{a|b})$$

$$\vec{\mu}_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$

$$= (0,0)^T + \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} \frac{1}{2\gamma^2 + \sigma^2} y = \left(\frac{\gamma^2(x_1 + x_2) + \gamma\epsilon}{2\gamma^2 + \sigma^2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)$$

$$\Sigma_{a|b} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \gamma \\ \gamma \end{pmatrix} \frac{1}{2\gamma^2 + \sigma^2} (\gamma, \gamma)$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{(2\gamma^2 + \sigma^2)} \begin{pmatrix} \gamma^2 & \gamma^2 \\ \gamma^2 & \gamma^2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\gamma^2 + \sigma^2}{2\gamma^2 + \sigma^2} & \frac{-\gamma^2}{2\gamma^2 + \sigma^2} \\ \frac{-\gamma^2}{2\gamma^2 + \sigma^2} & \frac{\gamma^2 + \sigma^2}{2\gamma^2 + \sigma^2} \end{pmatrix}$$

c.)

$$\rho = \frac{\frac{-\gamma^2}{2\gamma^2 + \sigma^2}}{\frac{\gamma^2 + \sigma^2}{2\gamma^2 + \sigma^2}} = \frac{-\gamma^2}{\gamma^2 + \sigma^2}$$

A