13.

Intro

In this question, we have to calculate the mutual information of two correlated gaussians r.v.'s. We should expect big mutual information values, when there is a linear relationship between the r.v's (i.e. $\rho = \pm 1$). On the other hand, we should expect small mutual information values, when the variables are decorrelated (i.e. $= \rho = 0$). **Before start the solution**, there is something that must be said. This question assumes that the reader have some knowledge which was not previously exposed in this chapter. The knowledge of how to calculate the distribution $p(x_1|x_2)$, wehre X_1 and X_2 are r.v.s with a bivariate normal distribution is only exposed in chapter 4 (Gaussian Models). Furthermore, this is not a straightfoward result that most people could do during a casual exercise like this. Thus, I will just state the result that we need in the beginning of the solution and ask the readers to believe it. Those more curious and familiar with linear algebra can see the actual proof in 4.3.4.

Solution

Let $x = (x_1, x_2)$ have a joint Gaussian distribution with parameters:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{1}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{2}$$

. Then, the conditional probability $p(x_1|x_2)$ is also Gaussian and its parameters are:

$$\begin{aligned} p(x_1|x_2) &= \mathcal{N}(x1|\mu_{1|2}, \Sigma_{1|2}) \\ \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned} \tag{3}$$

Hold this result for now and let's go to the actual question. The mutual information is given by:

$$I(X_1, X_2) = H(X_1) - H(X_1|X_2) \tag{4}$$

So, we need to calculate the two entropies. The first one is straightfoward, because $X \sim \mathcal{N}(0, \sigma^2)$ (those that don't know this result should also believe it or check chapter 4) and the author hinted how to calculate the entropy.

$$H(X) = \frac{1}{2} log_2[2\pi e \sigma^2] \tag{5}$$

The other is also straightfoward, since we know that $X_1|X_2$ is Gaussian and we know how to calculate the standard deviation. To calculate the standard deviation, we must first determine Σ for this distribtuion.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix} \tag{6}$$

From (6) and (3) we arrive at:

$$\begin{aligned}\Sigma_{1|2} = \sigma_{1|2}^2 = \sigma^2 - \rho\sigma^2 \frac{1}{\sigma^2} \rho\sigma^2 = \sigma^2(1-\rho^2) \\ \sigma_{1|2} = \sigma\sqrt{(1-\rho^2)}\end{aligned} \tag{7}$$

Knowing the standard deviation of $X_1|X_2$, we can calculate the conditional entropy:

$$H(X_1|X_2) = \frac{1}{2} log_2[2\pi e \sigma^2(1-\rho^2)] \tag{8}$$

Substituting (8) and (5) into (4), we get:

$$I(X_1, X_2) = H(X_1) - H(X_1|X_2) = \frac{1}{2} log_2[\frac{1}{(1-\rho^2)}] \tag{9}$$

For $\rho = \pm 1$:

$$I(X_1, X_2) = \frac{1}{2} log_2[\frac{1}{(1-1)}] = \infty \tag{10}$$

For $\rho = 0$:

$$I(X_1, X_2) = \frac{1}{2} log_2[\frac{1}{(1-0)}] = 0 \tag{11}$$

Conclusion As was said on the intro, the mutual information goes to larger value (infinity) when the two r.v's are linearly related, and it goes to 0 when there is no correlation between the variables. It is worthy to observe one key element of this result. The mutual information was introduced with the argument of being a more general measure of dependence between two variables than the correlation coeficient. However, in this scenario, the correlation coeficient is as powerful as the mutual infromation. Afterall, $\rho = 0$ iff $I(X, Y) = 0$. Therefore, if we have a bivariate normal distribution and $\rho = 0$, then X_1 and X_2 are independent.