1 Autoregressive Neural Networks

Consider, as the simplest non-trivial[1] example a three-qubit system acted upon by a single-qubit gate on the third qubit. New probabilities are give by

$$\begin{aligned} P'(a_1, a_2, a_3) &= \mathcal{O}^{a_3 a_2 a_1}_{b_3 b_2 b_1} \delta_{(a_2, b_2)} \delta_{(a_1, b_1)} P(b_1, b_2, b_3) \\ &= \mathcal{O}^{a_3}_{b_3} P(a_1, a_2, b_3) \end{aligned}$$

The autoregressive property ensures that the conditional probabilities, in which the joint probability can be decomposed, are implemented by seperate neural networks. In probabilities, the following conditions hold

$$\begin{aligned} P(a_1, a_2, a_3) &= P(a_1) P(a_2|a_1) P(a_3|a_2 a_1), \\ P'(a_1, a_2, a_3) &= P(a_1) P(a_2|a_1) P'(a_3|a_2 a_1). \end{aligned}$$

Hence we can write

$$P'(a_3|a_2 a_1) = \mathcal{O}^{a_3}_{b_3} P(b_3|a_2 a_1) \tag{1}$$

Assuming the conditional probabilites for the measurement outcome on qubit three are encooded by 1-layer feed-forward networks with softmax activation function allows us to rewrite 1 as

$$\frac{\exp\left(\beta'^{3}_{a_3} + \gamma'^{(23)}_{a_3 a_2} + \gamma'^{(13)}_{a_3 a_1}\right)}{\sum_{i,j,k} \exp\left(\beta'^{3}_{i} \gamma'^{(23)}_{ij} + \gamma'^{(13)}_{i,}\right)} = \sum_{b_3} \mathcal{O}^{a_3}_{b_3} \frac{\exp\left(\beta'^{3}_{b_3} + \gamma'^{(23)}_{b_3 a_2} + \gamma'^{(13)}_{b_3 a_1}\right)}{\sum_{i,j,k} \exp\left(\beta'^{3}_{i} \gamma'^{(23)}_{ij} + \gamma'^{(13)}_{i,}\right)}$$

[1] not overparametrized