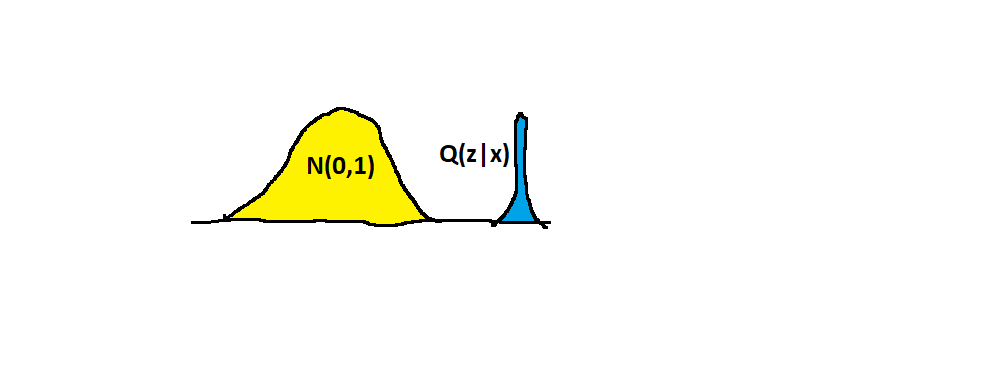
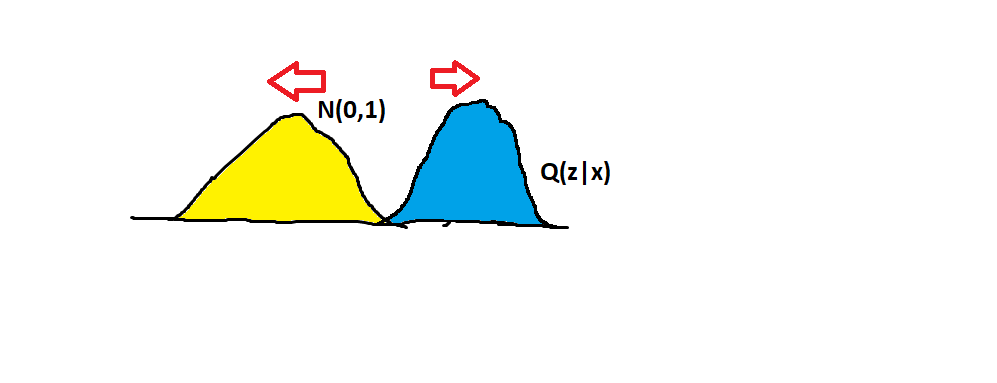
* Data fidelity 🗹 K.L divergence 🗷

Without regularization, network will cheat by learning narrow distribution. With very narrow variance the distribution effectively representing single value like in auto encoder. 

* Data fidelity 🗷 K.L divergence 🗹

Penalizing the reconstruction loss encourages our distribution to deviate from the prior to describe some characteristics of data.



* Data fidelity 🗹 K.L divergence 🗹

1st term results in the attraction between the two distribution. 2nd term ensures the sufficient variance in the distribution so that the model can produce samples of varying attributes.

