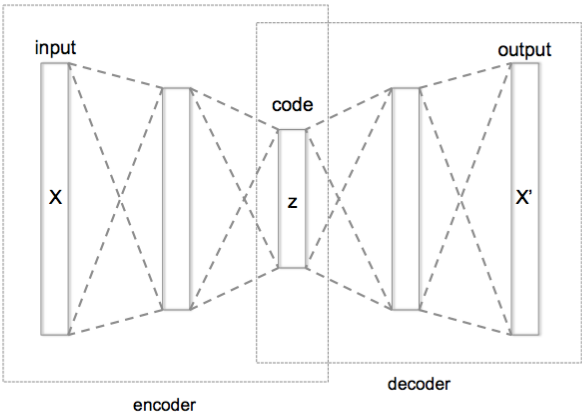
Rappel théorique sur les VAE

A) Mécanisme encodeur/décodeur

Les structures encodeur/décodeur sont utilisées dans des situations où l'objectif est de réduire la dimension d'un problème, par exemple pour obtenir une représentation plus compacte du domaine des images de chiffres en noir et blanc (domaine MNIST).



https://fr.wikipedia.org/wiki/Auto-encodeur

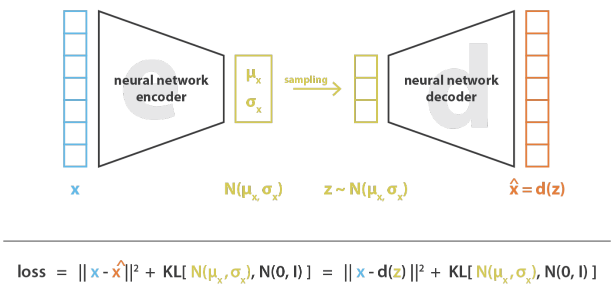
L'idée est de construire un réseau de neurones en forme de « sablier » : à partir d'une entrée X on construira une représentation encodée $e(X) = z$ de dimension inférieure. À partir de cette représentation encodée on doit être capable de retrouver un élément proche de X noté X'. Il s'agit en fait de constituer un « mapping » entre l'espace des X et un espace latent de plus petite dimension. Pour entraîner un tel réseau on contrôlera une fonction de perte dite de « reconstruction » de la forme $|X - X'|^2$.

B) Utilisation d'un mécanisme encodeur/décodeur à des fins d'échantillonage

À partir de l'encodage obtenu, on espère pouvoir échantillonner dans l'espace latent, puis appliquer le décodeur afin d'obtenir des éléments représentatifs de l'ensemble de définition de X. Mais pour que cette stratégie soit efficace il importe que l'espace latent soit suffisamment **régularisé**.
La régularité de l'espace latent peut être comprise comme la combinaison de sa continuité (deux vecteurs latents proches doivent avoir des représentations décodées proches) et de son exhaustivité (aucun vecteur latent ne doit produire de représentation aberrante une fois décodé).

C) Introduction d'une distribution et VAE

L'idée clé des VAE est de décrire l'espace latent par une distribution régularisée. L'encodeur donnera en sortie les paramètres μ_X et σ_X d'une loi normale. Chacun des paramètres (μ_i, σ_i) de ces vecteurs permettront de simuler le i-ème élément d'entrée du décodeur, avec i variant entre 1 et m (taille de l'espace latent)).



https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Il convient de modifier légèrement la fonction de perte. La nouvelle fonction de perte peut être divisée en deux parties :

- un terme de **reconstruction** (semblable à la fonction de perte utilisée pour un autoencodeur classique)
- un terme de **régularisation** qui permettra de vérifier que notre distribution intermédiaire se rapproche d'une $\mathcal{N}(0,1)$.

Ainsi la nouvelle perte s'écrit : $\|X - X'\| + KL[\mathcal{N}(\mu_X, \sigma_X), \mathcal{N}(0,1)]$.