Autoencoder regularizado

codificador $= X \rightarrow Z$

Decodificador $= Z \rightarrow \hat{X}$

Modelo:

$$Z = f(X; W_1, b_1) = \phi(W_1 X + b_1)$$

W_1 = matriz de pesos del codificador

b_1 = es el vector de sesgos del codificador

ϕ = función de activación

Z = Espacio latente

$$\hat{X} = g(Z; W_2, b_2) = \psi(W_2 Z + b_2)$$

W_2 - Matriz de pesos del decodificador

b_2 = vector de sesgos del decodificador

ψ = función de activación

$\hat{X}$ = Reconstrucción de la entrada

Función de costo:

$$E_X \{ L(X, \hat{X}) \} + \lambda(w)$$

λ = Regularizador

Función de optimización

$$\hat{\Theta} = \arg\min \mathbb{E}_x \{ \mathcal{L}(x, \hat{x}) + \lambda(w) \}$$

$$\hat{\Theta} = \arg\min \sum_{N} \sum_{D} x_{ND} \log(\hat{x}_{ND}) + \lambda \|w\|_2^2$$

Autoencoder Variacional

Modelo:

$$z = f(x; W_1, b_1) = \phi(W_1 x + b_1)$$

$$\hat{x} = g(z; W_2, b_2) = \psi(W_2 z + b_2)$$

$P(z | x, \Theta)$ Prior $\qquad q_\theta(z | x)$ Posterior

$$p(z) \sim N(z; 0 I)$$

En el encoder $\Rightarrow$ $q_\theta(z | x) - N(z; \mu(x), \sigma^2(x) I)$

$$\Rightarrow z = \mu(x) + \sigma(x) \odot N(0, I)$$

En el decoder: $P(x | z) = N(x; \mu(z), \sigma^2(z) I)$

Función de costo

$$KL(q(z | x; \theta) \| p(z | x; \theta)) + \mathcal{L}(x, \hat{x})$$

$$= KL(q(z|x)\|p(z)) - E_q\{\log p(x|z)\} + \log p(\hat{x}) + \mathcal{L}(x,\hat{x})$$

Función de optimización

$$\hat{\Theta} = \arg\min_{\Theta} \mathcal{L}(x,\hat{x}) + KL(q(z|x,\Theta)\|p(z)) - E_q\{\log p(x|z)\}$$

Redes generativas adversarias (GAN)

Modelo

Generador $x \in \mathbb{R}^{D}$

$\hat{x} \in \mathbb{R}^{D \times D_1}$

$\Rightarrow \hat{x} = \phi(W_1 x + b_2)$ $\quad \Theta_1 = W_1, b_1$

Aquí se espera que $\hat{x}$ siga un prior $p_h(H)$

Discriminador $\hat{x} \in \mathbb{R}^{D \times D_1}$

$y \in [0,1]$

$$\Rightarrow y = \Psi(\hat{x}, \Theta_2) = \Psi(W_2 \hat{x} + b_2)$$

$$y = \begin{cases} 1 & \text{si } x \in X \\ 0 & \text{si } x \in \hat{X} \end{cases}$$

Función de costo

$$\mathcal{L}_g = -E_{H \sim p(H)}\{\log y(\hat{x}(H))\}$$

$$\mathcal{L}_d = -\mathbb{E}_{x \sim real}\{\log(\gamma(x))\} - \mathbb{E}_{h \sim p(H)}\{1 - \log \gamma(\hat{x}(x))\}$$

Función de optimización

$$\hat{\theta}_1, \hat{\theta}_2 = \min_{\theta_1} \max \mathcal{L}_d(\theta_2) - \mathcal{L}_g(\theta_1)$$