

Generación de música mediante redes neuronales profundas

Trabajo de fin de grado

Antonio Martín Ruiz

Universidad de Granada

17 de septiembre de 2020

Curso 2019-2020

Índice

Aprendizaje profundo

Aproximación por superposición de funciones sigmoidales

Tratamiento de secuencias

Aprendizaje de características

MusicVAE

AutoLoops

Redes neuronales profundas

$$f^*(\mathbf{x}) \approx f(\mathbf{x}; \theta) = (f_1 \circ f_2 \circ \dots \circ f_N)(\mathbf{x})$$

$$f_i(\mathbf{x}; \mathbf{W}_i, \mathbf{b}_i) = \sigma(\mathbf{W}_i^T \mathbf{x} + \mathbf{b}_i)$$

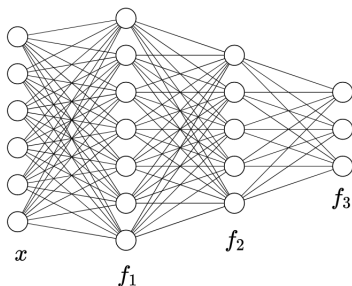


Figura: Ejemplo de red neuronal prealimentada con dos capas ocultas

Teorema de Hahn Banach

Teorema de extensión de Hahn-Banach. X espacio normado, M subespacio de X , $g \in M^*$ entonces $\exists f \in X^*$ que extiende a g con $\|f\| = \|g\|$.

Teorema de Hahn Banach

Teorema de extensión de Hahn-Banach. X espacio normado, M subespacio de X , $g \in M^*$ entonces $\exists f \in X^*$ que extiende a g con $\|f\| = \|g\|$.

Corolario. X espacio normado, M un subespacio de X . Si $x_0 \notin \overline{M}$, $\exists f \in X^*$ tal que $f(x) = 0 \forall x \in M$, $f(x_0) = 1$, y $\|f\| = \frac{1}{d}$, donde d es la distancia de x_0 a M .

Teorema de Hahn Banach

Teorema de extensión de Hahn-Banach. X espacio normado, M subespacio de X , $g \in M^*$ entonces $\exists f \in X^*$ que extiende a g con $\|f\| = \|g\|$.

Corolario. X espacio normado, M un subespacio de X . Si $x_0 \notin \overline{M}$, $\exists f \in X^*$ tal que $f(x) = 0 \forall x \in M$, $f(x_0) = 1$, y $\|f\| = \frac{1}{d}$, donde d es la distancia de x_0 a M .

Demostración. $\gamma = x + ax_0$. $f(x + ax_0) = a$ lineal, $\|f\| = \frac{1}{d}$. Aplicar extensión.

Teorema de representación de Riesz

Teorema. X un espacio de Hausdorff localmente compacto, T funcional lineal acotado sobre $C_0(X)$. $\exists \mathcal{M}$ σ -álgebra en X que contiene todos los conjuntos de Borel en X y $\exists! \mu$ medida con signo regular sobre \mathcal{M} tal que

$$T(f) = \int_X f d\mu$$

para cada $f \in C_0(X)$. Además

$$\|T\| = |\mu|(X).$$

Teorema para funciones discriminatorias

Definición. $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ es *discriminatoria* si, cuando para una medida $\mu \in M(I^n)$ se tiene que

$$\int_{I^n} \sigma(\mathbf{y}^T \mathbf{x} + \theta) d\mu(\mathbf{x}) = 0$$

para todo $\mathbf{y} \in \mathbb{R}^n$, $\theta \in \mathbb{R}$ implica que $\mu = 0$.

Teorema para funciones discriminatorias

Definición. $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ es *discriminatoria* si, cuando para una medida $\mu \in M(I^n)$ se tiene que

$$\int_{I^n} \sigma(\mathbf{y}^T \mathbf{x} + \theta) d\mu(\mathbf{x}) = 0$$

para todo $\mathbf{y} \in \mathbb{R}^n$, $\theta \in \mathbb{R}$ implica que $\mu = 0$.

Teorema. σ continua discriminatoria. Entonces

$$S = \left\{ g : g(\mathbf{x}) = \sum_{j=1}^k \alpha_j \sigma(\mathbf{y}_j^T \mathbf{x} + \theta_j) \right\}$$

es denso en $C(I^n)$.

Teorema para funciones discriminatorias

Definición. $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ es *discriminatoria* si, cuando para una medida $\mu \in M(I^n)$ se tiene que

$$\int_{I^n} \sigma(\mathbf{y}^T \mathbf{x} + \theta) d\mu(\mathbf{x}) = 0$$

para todo $\mathbf{y} \in \mathbb{R}^n$, $\theta \in \mathbb{R}$ implica que $\mu = 0$.

Teorema. σ continua discriminatoria. Entonces

$$S = \left\{ g : g(\mathbf{x}) = \sum_{j=1}^k \alpha_j \sigma(\mathbf{y}_j^T \mathbf{x} + \theta_j) \right\}$$

es denso en $C(I^n)$.

Demostración. Si $\overline{S} \neq C(I^n)$, aplicar corolario de H-B para obtener funcional, que por Riezs es $F(h) = \int_{I^n} h(\mathbf{x}) d\mu(\mathbf{x})$, nulo.

Contradicción.

Lema de funciones sigmoidales

Lema. Las funciones sigmoidales continuas son discriminatorias.

Lema de funciones sigmoidales

Lema. Las funciones sigmoidales continuas son discriminatorias.

Demostración.

$$\sigma(\lambda(\mathbf{y}^T \mathbf{x} + \theta) + \phi) \begin{cases} \rightarrow 1 & \text{si } \mathbf{y}^T \mathbf{x} + \theta > 0 \text{ cuando } \lambda \rightarrow +\infty, \\ \rightarrow 0 & \text{si } \mathbf{y}^T \mathbf{x} + \theta < 0 \text{ cuando } \lambda \rightarrow +\infty, \\ = \sigma(\phi) & \text{si } \mathbf{y}^T \mathbf{x} + \theta = 0 \text{ para todo } \lambda. \end{cases}$$

converge puntualmente a

$$\gamma(\mathbf{x}) = \begin{cases} 1 & \text{si } \mathbf{y}^T \mathbf{x} + \theta > 0 \\ 0 & \text{si } \mathbf{y}^T \mathbf{x} + \theta < 0 \\ \sigma(\phi) & \text{si } \mathbf{y}^T \mathbf{x} + \theta = 0 \end{cases}$$

Por convergencia dominada,

$$0 = \lim_{\lambda \rightarrow \infty} \int_{I^n} \sigma(\lambda(\mathbf{y}^T \mathbf{x} + \theta) + \phi) d\mu(\mathbf{x}) = \sigma(\phi) \mu(\Pi_{\mathbf{y}, \theta}) + \mu(H_{\mathbf{y}, \theta})$$

$$F(h) = \int_{I^n} h(\mathbf{y}^T \mathbf{x}) d\mu(\mathbf{x}), h \in L^\infty(\mathbb{R})$$

Tomando h función indicadora en $[\theta, +\infty)$.

$$F(h) = \int_{I^n} h(\mathbf{y}^T \mathbf{x}) d\mu(\mathbf{x}) = \mu(\Pi_{\mathbf{y}, -\theta}) + \mu(H_{\mathbf{y}, -\theta}) = 0 \Rightarrow F = 0$$

Tomando $s(u) = \sin(m \cdot u)$, $c(u) = \cos(m \cdot u)$,

$$\begin{aligned} F(s+ic) &= \int_{I^n} (\cos(m^T x) + i \sin(m^T x)) d\mu(x) = \int_{I^n} \exp(im^T x) d\mu(x) = 0 \\ &\Rightarrow \mu = 0 \end{aligned}$$

Teorema para funciones sigmoidales

Teorema. σ sigmoideal continua. Entonces

$$S = \left\{ g : g(\mathbf{x}) = \sum_{j=1}^k \alpha_j \sigma(\mathbf{y}_j^T \mathbf{x} + \theta_j) \right\}$$

es denso en $C(I^n)$.

Otros resultados

- Generalización para funciones acotadas no constantes.
- Generalización para funciones Riemann-integrables no polinomiales.
- Anchura fija y profundidad arbitraria.

Redes neuronales recurrentes

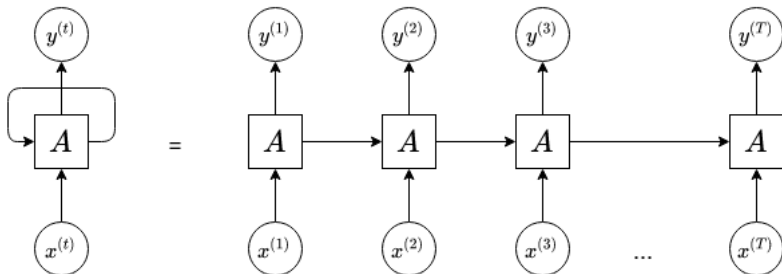


Figura: Desenrollado de una red recurrente

Redes recurrentes bidireccionales y profundas

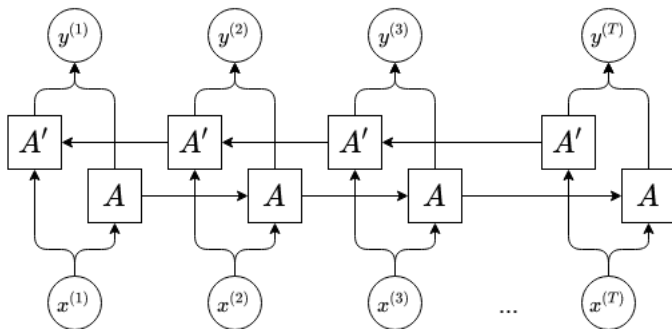


Figura: Red recurrente bidireccional

Redes recurrentes con puertas

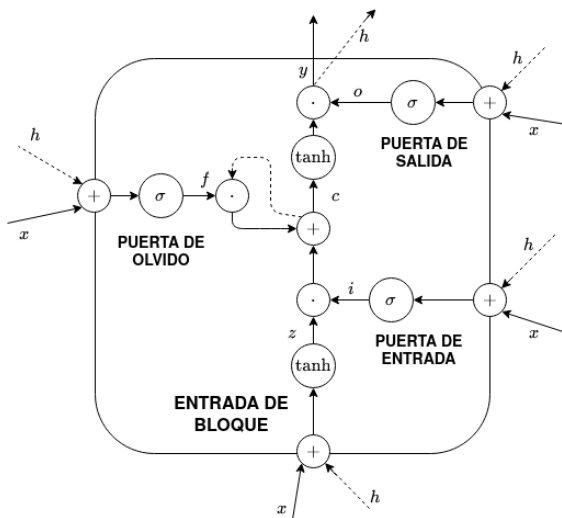


Figura: Bloque LSTM (Long short-term memory)

Autoencoder

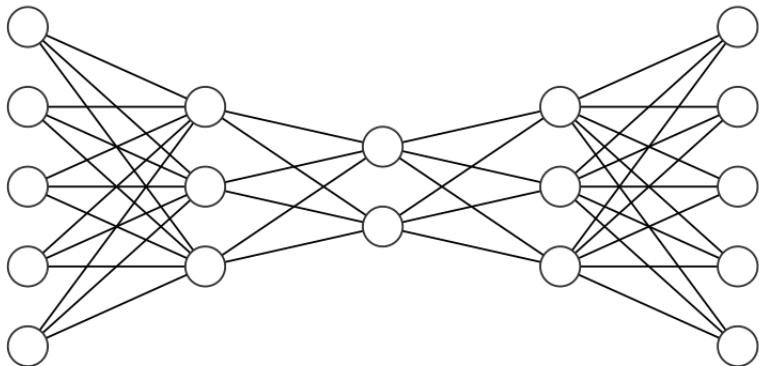


Figura: Estructura de un *autoencoder* como red prealimentada profunda

Autoencoder variacional

$$p_{\theta}(\mathbf{z}|\mathbf{x}) = \frac{p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})}{p_{\theta}(\mathbf{x})}$$

$$p_{\theta}(\mathbf{z}|\mathbf{x}) \approx q_{\phi}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mu, \sigma^2 \mathbf{I})$$

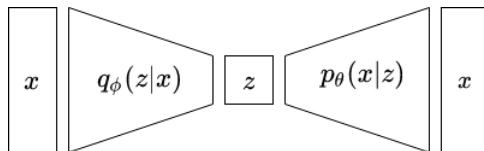


Figura: Estructura de un *autoencoder* variacional

Función de coste del VAE

Definición. P y Q distribuciones sobre el mismo espacio de probabilidad, se define la divergencia de Kullback-Leibler como

$$D_{KL}(Q \parallel P) = E_Q \left[\log \frac{Q(x)}{P(x)} \right].$$

Proposición. Dadas las distribuciones de probabilidad $p(\mathbf{x}) = N(\mathbf{x}; \mu_1, \Sigma_1)$ y $q(\mathbf{x}) = N(\mathbf{x}; \mu_2, \Sigma_2)$, ambas de dimensión k , se cumple que $D_{KL}(p(\mathbf{x}) \parallel q(\mathbf{x})) =$

$$\frac{1}{2} \left(\log \frac{|\Sigma_1|}{|\Sigma_2|} - k + \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) \right).$$

Cálculo de ELBO (Evidente lower bound)

$$D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p_{\theta}(\mathbf{z}|\mathbf{x})) = \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{z}|\mathbf{x})] =$$

Cálculo de ELBO (Evidente lower bound)

$$\begin{aligned} D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p_\theta(\mathbf{z}|\mathbf{x})) &= \\ E_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} [\log q_\phi(\mathbf{z}|\mathbf{x}) - \log p_\theta(\mathbf{z}|\mathbf{x})] &= \\ E_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} \left[\log q_\phi(\mathbf{z}|\mathbf{x}) - \log \frac{p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})}{p_\theta(\mathbf{x})} \right] &= \\ E_{\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})} [\log q_\phi(\mathbf{z}|\mathbf{x}) - \log p_\theta(\mathbf{x}|\mathbf{z}) - \log p_\theta(\mathbf{z}) + \log p_\theta(\mathbf{x})] . \end{aligned}$$

Cálculo de ELBO (Evidente lower bound)

$$\begin{aligned} D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p_{\theta}(\mathbf{z}|\mathbf{x})) &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{z}|\mathbf{x})] &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \left[\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log \frac{p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})}{p_{\theta}(\mathbf{x})} \right] &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{x}|\mathbf{z}) - \log p_{\theta}(\mathbf{z}) + \log p_{\theta}(\mathbf{x})] . \end{aligned}$$

$$\begin{aligned} D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p_{\theta}(\mathbf{z}|\mathbf{x})) - \log p_{\theta}(\mathbf{x}) &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{x}|\mathbf{z}) - \log p_{\theta}(\mathbf{z})] &= \end{aligned}$$

Cálculo de ELBO (Evidente lower bound)

$$\begin{aligned} D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p_{\theta}(\mathbf{z}|\mathbf{x})) &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{z}|\mathbf{x})] &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} \left[\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log \frac{p_{\theta}(\mathbf{x}|\mathbf{z})p_{\theta}(\mathbf{z})}{p_{\theta}(\mathbf{x})} \right] &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{x}|\mathbf{z}) - \log p_{\theta}(\mathbf{z}) + \log p_{\theta}(\mathbf{x})] & . \end{aligned}$$

$$\begin{aligned} D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p_{\theta}(\mathbf{z}|\mathbf{x})) - \log p_{\theta}(\mathbf{x}) &= \\ E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{x}|\mathbf{z}) - \log p_{\theta}(\mathbf{z})] &= \\ -E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z})] + E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log q_{\phi}(\mathbf{z}|\mathbf{x}) - \log p_{\theta}(\mathbf{z})] &= \\ -E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z})] + D_{KL}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p_{\theta}(\mathbf{z})) & . \end{aligned}$$

Truco de reparametrización

$$L(\theta, \phi; \mathbf{x}) = -E_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z})] + \frac{1}{2} \sum_{j=1}^k (-\log(\sigma_j^2) - 1 + \sigma_j^2 + \mu_j^2) .$$

$$\tilde{L}(\theta, \phi; \mathbf{x}) = -\frac{1}{\mathcal{L}} \sum_{l=1}^{\mathcal{L}} \log p_{\theta}(\mathbf{x}^{(i)} | \mathbf{z}^{(i,l)}) + \frac{1}{2} \sum_{j=1}^k (-\log(\sigma_j^2) - 1 + \sigma_j^2 + \mu_j^2) .$$

MusicVAE

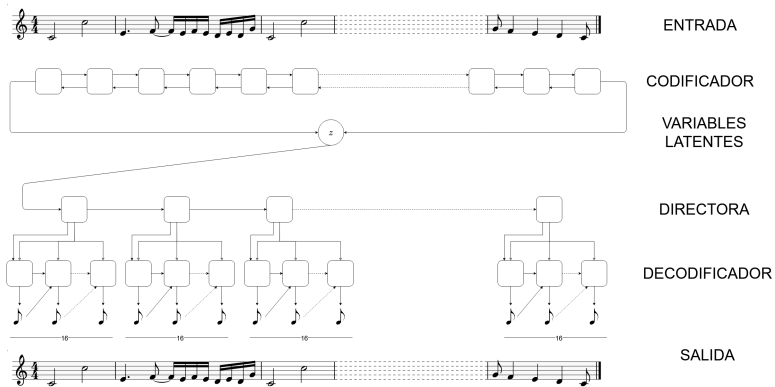


Figura: Estructura del modelo MusicVAE

AutoLoops