Practica 4:
Modelos de Difusión

Ana Paula Morresi
Aprendizaje profundo con aplicación a visión artificial

26 de noviembre de 2025

1. Problemas

Ejercicio 1.

Proceso Directo

El proceso iterativo (proceso directo) está definido por:

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon_t \quad (1)$$

Donde $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\alpha_t = 1 - \beta_t$, *y* $\tilde{\alpha}_t = \prod_{k=1}^{t} \alpha_k$.

Deduciremos la distribución condicional $x_t|x_0$.

a) Distribución de probabilidad de x_t para $t > 0$

Dado que ϵ_t tiene un distribución gaussiana, x_t también tendrá distribución gaussiana. Haciendo una expansión recursiva podemos obtener x_t en función de x_0. Para esto, sustituimos la definición de x_{t-1} en la ecuación (1), y luego repetimos la sustitución hasta llegar al estado inicial x_0:

$$\begin{aligned}
x_t &= \sqrt{\alpha_t}x_{t-1} + \sqrt{\beta_t}\epsilon_t \\
x_t &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{\beta_{t-1}}\epsilon_{t-1}\right) + \sqrt{\beta_t}\epsilon_t \\
x_t &= \sqrt{\alpha_t\alpha_{t-1}}x_{t-2} + \sqrt{\alpha_t\beta_{t-1}}\epsilon_{t-1} + \sqrt{\beta_t}\epsilon_t \\
&\vdots \\
x_t &= \left(\prod_{k=1}^{t}\sqrt{\alpha_k}\right)x_0 + \sum_{k=1}^{t-1}\left(\sqrt{\beta_k}\prod_{j=k+1}^{t}\sqrt{\alpha_j}\right)\epsilon_k + \sqrt{\beta_t}\epsilon_t
\end{aligned}$$

Usando la notación $\tilde{\alpha}_t$:

$$x_t = \sqrt{\tilde{\alpha}_t}x_0 + \sum_{k=1}^{t}\left(\sqrt{\beta_k}\sqrt{\frac{\tilde{\alpha}_t}{\tilde{\alpha}_k}}\right)\epsilon_k$$

Es claro que x_t es una combinación lineal de variables gaussianas independientes (ϵ_k), más un término constante ($\sqrt{\tilde{\alpha}_t}x_0$). Por lo tanto, x_t es una variable aleatoria gaussiana.

b) Cálculo de la Media $\mathbb{E}[x_t]$

La esperanza de x_t es:

$$\begin{aligned}
\mathbb{E}[x_t|x_0] &= \mathbb{E}\left[\sqrt{\tilde{\alpha}_t}x_0 + \sum_{k=1}^{t}\left(\sqrt{\beta_k}\prod_{j=k+1}^{t}\sqrt{\alpha_j}\right)\epsilon_k\right] \\
&= \sqrt{\tilde{\alpha}_t}x_0 + \sum_{k=1}^{t}\left(\sqrt{\beta_k}\prod_{j=k+1}^{t}\sqrt{\alpha_j}\right)\mathbb{E}[\epsilon_k]
\end{aligned}$$

Dado que $\mathbb{E}[\epsilon_k] = \mathbf{0}$ para todo k:

$$\mathbb{E}[x_t|x_0] = \sqrt{\tilde{\alpha}_t}x_0$$

c) Cálculo de la Covarianza Cov$[x_t]$

La matriz de covarianza de x_t es la covarianza del término de ruido, ya que x_0 es fijo. Debido a la independencia de ϵ_k, $\mathrm{Cov}[\sum a_k \epsilon_k] = \sum a_k^2 \mathrm{Cov}[\epsilon_k]$:

$$
\begin{aligned}
\mathrm{Cov}[x_t|x_0] &= \sum_{k=1}^{t} \mathrm{Cov}\left[\left(\sqrt{\beta_k} \prod_{j=k+1}^{t} \sqrt{\alpha_j}\right) \epsilon_k\right] \\
&= \sum_{k=1}^{t} \left(\sqrt{\beta_k} \prod_{j=k+1}^{t} \sqrt{\alpha_j}\right)^2 \mathrm{Cov}[\epsilon_k] \\
&= \sum_{k=1}^{t} \left(\beta_k \prod_{j=k+1}^{t} \alpha_j\right) \mathbf{I}
\end{aligned}
$$

Luego se tiene, usando la serie telescópica $\sum_{n=1}^{N}(a_n - a_{n+1}) = a_1 - a_{N+1}$, que:

$$
\begin{aligned}
\sum_{k=1}^{t} \left(\beta_k \prod_{j=k+1}^{t} \alpha_j\right) &= \sum_{k=1}^{t-1} \left(\beta_k \prod_{j=k+1}^{t} \alpha_j\right) + \beta_t \\
&= \sum_{k=1}^{t-1} \left((1-\alpha_k) \prod_{j=k+1}^{t} \alpha_j\right) + \beta_t \\
&= \sum_{k=1}^{t-1} \left(\prod_{j=k+1}^{t} \alpha_j - \alpha_k \prod_{j=k+1}^{t} \alpha_j\right) + \beta_t \\
&= \sum_{k=1}^{t-1} \left(\prod_{j=k+1}^{t} \alpha_j - \prod_{j=k}^{t} \alpha_j\right) + \beta_t \\
&= \alpha_t - \prod_{j=1}^{t} \alpha_j + (1-\alpha_t) = 1 - \prod_{k=1}^{t} \alpha_k = 1 - \tilde{\alpha}_t
\end{aligned}
$$

De esta forma la covarianza queda:

$$
\mathrm{Cov}[x_t|x_0] = (1 - \tilde{\alpha}_t)\,\mathbf{I}
$$

Entonces, la distribución de probabilidad de x_t para $t > 0$, condicionada a la muestra inicial x_0, es una distribución gaussiana con los parámetros demostrados:

$$
x_t|x_0 \sim \mathcal{N}\left(\sqrt{\tilde{\alpha}_t} x_0, (1-\tilde{\alpha}_t)\mathbf{I}\right)
$$

Con esto es fácil ver que la distribución de probabilidad de x_t condicionada a x_{t-1}, es también es una distribución gaussiana:

$$
x_t|x_{t-1} \sim \mathcal{N}\left(\sqrt{\alpha_t} x_{t-1}, \beta_t \mathbf{I}\right)
$$

Ejercicio 2.

Proceso Inverso

El proceso inverso se define por la relación de recurrencia:

$$x_{t-1} = \mu(x_t) + \sqrt{\beta_t}\epsilon_t \quad (2)$$

O condicionado a x_0:

$$x_{t-1} = \mu(x_t, x_0) + \sqrt{\beta_t}\epsilon_t \quad (3)$$

Esta relación define la distribución condicional $p(x_{t-1}|x_t, x_0)$.

a) Cálculo de $p(x_{t-1}|x_t, x_0)$ usando la Regla de Bayes

Usando la Regla de Bayes, $p(A|B) = \frac{p(B|A)p(A)}{p(B)}$, aplicada a $p(x_{t-1}|x_t, x_0)$:

$$p(x_{t-1}|x_t, x_0) = \frac{p(x_t|x_{t-1}, x_0)p(x_{t-1}|x_0)}{p(x_t|x_0)}$$

b) Calculo de la Media del Proceso Inverso Condicionado $\mu(x_t, x_0)$

Utilizando las distribuciones gaussianas del proceso directo (obtenidas en la pregunta 1):

$$p(x_t|x_0) = \mathcal{N}\left(x_t; \sqrt{\tilde{\alpha}_t}x_0, (1-\tilde{\alpha}_t)\mathbf{I}\right)$$

$$p(x_{t-1}|x_0) = \mathcal{N}\left(x_{t-1}; \sqrt{\tilde{\alpha}_{t-1}}x_0, (1-\tilde{\alpha}_{t-1})\mathbf{I}\right)$$

$$p(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{\alpha_t}x_{t-1}, \beta_t\mathbf{I}\right)$$

La distribución $p(x_{t-1}|x_t, x_0)$ resulta ser también gaussiana, $p(x) \propto e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$. Para encontrar su media $\mu(x_t, x_0)$, examinamos el producto de las densidades de probabilidad.

Aunque $p(x_t|x_0)$ es una de las tres distribuciones gaussianas del proceso directo, esta una constante de normalización (respecto a x_{t-1}). Cuando realizamos la multiplicación de las densidades para encontrar la forma de $p(x_{t-1}|x_t, x_0)$, solo nos interesa cómo la densidad varía con respecto a la variable que estamos condicionando, que es $\mathbf{x_{t-1}}$. La distribución $p(x_t|x_0)$ es una función de x_t y x_0. Por lo tanto, según lo que vimos antes nos interesa:

$$p(x_{t-1}|x_t, x_0) \propto p(x_t|x_{t-1})p(x_{t-1}|x_0)$$

Dado que las distribuciones son gaussianas, el producto de las densidades es proporcional a una nueva densidad gaussiana:

$$p(x_{t-1}|x_t, x_0) \sim \mathcal{N}(x_{t-1}; \mu(x_t, x_0), \tilde{\beta}_t\mathbf{I})$$

Para una distribución gaussiana con varianza $\Sigma = \sigma^2\mathbf{I}$, el logaritmo de su densidad es, ignorando términos constantes (que no dependen de la variable x):

$$\log p(x) \propto -\frac{1}{2\sigma^2}\|x-\mu\|^2$$

Expandiendo el logaritmo del producto:

$$\begin{aligned}\log p(x_{t-1}|x_t, x_0) &\propto \log p(x_t|x_{t-1}) + \log p(x_{t-1}|x_0)\\ &\propto -\frac{1}{2\beta_t}\|x_t - \sqrt{\alpha_t}x_{t-1}\|^2 - \frac{1}{2(1-\tilde{\alpha}_{t-1})}\|x_{t-1} - \sqrt{\tilde{\alpha}_{t-1}}x_0\|^2\end{aligned}$$

Expandiendo los términos cuadráticos y agrupando los coeficientes para x_{t-1}:

$$\log p \propto -\frac{1}{2}x_{t-1}^T \underbrace{\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\tilde{\alpha}_{t-1}}\right)}_{\text{Coeficiente } x_{t-1}^2 \implies \tilde{\beta}_t^{-1}} x_{t-1} + x_{t-1}^T \underbrace{\left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\tilde{\alpha}_{t-1}}}{1-\tilde{\alpha}_{t-1}}x_0\right)}_{\text{Coeficiente } x_{t-1} \implies \tilde{\beta}_t^{-1}\mu(x_t,x_0)}$$

Ahora vamos a calcular la varianza inversa $\tilde{\beta}_t^{-1}$. Para esto definimos la inversa de la varianza $\tilde{\beta}_t^{-1}$ a partir del coeficiente cuadrático:

$$\tilde{\beta}_t^{-1} = \frac{\alpha_t}{\beta_t} + \frac{1}{1-\tilde{\alpha}_{t-1}}$$

Sumamos las fracciones:

$$\tilde{\beta}_t^{-1} = \frac{\alpha_t(1-\tilde{\alpha}_{t-1}) + \beta_t}{\beta_t(1-\tilde{\alpha}_{t-1})}$$

Usando $\alpha_t + \beta_t = 1$:

$$\tilde{\beta}_t^{-1} = \frac{(\alpha_t + \beta_t) - \alpha_t\tilde{\alpha}_{t-1}}{\beta_t(1-\tilde{\alpha}_{t-1})} = \frac{1 - \alpha_t\tilde{\alpha}_{t-1}}{\beta_t(1-\tilde{\alpha}_{t-1})}$$

Usando $\tilde{\alpha}_t = \alpha_t\tilde{\alpha}_{t-1}$:

$$\tilde{\beta}_t^{-1} = \frac{1-\tilde{\alpha}_t}{\beta_t(1-\tilde{\alpha}_{t-1})} \implies \tilde{\beta}_t = \frac{\beta_t(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t}$$

De esta forma podemos calcular la media $\mu(x_t, x_0)$. Esta se obtiene a partir del coeficiente lineal, multiplicando por $\tilde{\beta}_t$:

$$\mu(x_t, x_0) = \tilde{\beta}_t\left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\tilde{\alpha}_{t-1}}}{1-\tilde{\alpha}_{t-1}}x_0\right)$$

Sustituimos la expresión obtenida para $\tilde{\beta}_t$:

$$\mu(x_t, x_0) = \frac{\beta_t(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t}\left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\tilde{\alpha}_{t-1}}}{1-\tilde{\alpha}_{t-1}}x_0\right)$$

Distribuimos el factor $\tilde{\beta}_t$ a cada término:

1. **Término de x_t:**

$$\frac{\beta_t(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t} \cdot \frac{\sqrt{\alpha_t}}{\beta_t}x_t = \frac{\sqrt{\alpha_t}(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t}x_t$$

2. **Término de x_0:**

$$\frac{\beta_t(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t} \cdot \frac{\sqrt{\tilde{\alpha}_{t-1}}}{1-\tilde{\alpha}_{t-1}}x_0 = \frac{\beta_t\sqrt{\tilde{\alpha}_{t-1}}}{1-\tilde{\alpha}_t}x_0$$

Luego, sumando los términos:

$$\mu(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t}x_t + \frac{\sqrt{\tilde{\alpha}_{t-1}}\beta_t}{1-\tilde{\alpha}_t}x_0 \quad (4)$$

que es lo que estábamos buscando.

c) Eliminación de x_0 para obtener $\mu(x_t)$

Vamos a reescribir $\mu(x_t, x_0)$ en función de x_t y el ruido ϵ_t para obtener $\mu(x_t)$. Para esto, partimos de la ecuación del proceso directo que relaciona x_t y x_0 con el ruido agregado ϵ_t:

$$x_t = \sqrt{\tilde{\alpha}_t} x_0 + \sqrt{1 - \tilde{\alpha}_t} \epsilon_t \quad (5)$$

donde $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Despejando x_0:

$$x_0 = \frac{1}{\sqrt{\tilde{\alpha}_t}} \left(x_t - \sqrt{1 - \tilde{\alpha}_t} \epsilon_t \right)$$

Sustituimos esta expresión para x_0 en la ecuación (4):

$$\begin{aligned}
\mu(x_t, x_0) &= \frac{\sqrt{\alpha_t}(1 - \tilde{\alpha}_{t-1})}{1 - \tilde{\alpha}_t} x_t + \frac{\sqrt{\tilde{\alpha}_{t-1}} \beta_t}{1 - \tilde{\alpha}_t} \left[\frac{1}{\sqrt{\tilde{\alpha}_t}} \left(x_t - \sqrt{1 - \tilde{\alpha}_t} \epsilon_t \right) \right] \\
&= x_t \left[\frac{\sqrt{\alpha_t}(1 - \tilde{\alpha}_{t-1})}{1 - \tilde{\alpha}_t} + \frac{\sqrt{\tilde{\alpha}_{t-1}} \beta_t}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} \right] - \frac{\sqrt{\tilde{\alpha}_{t-1}} \beta_t \sqrt{1 - \tilde{\alpha}_t}}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} \epsilon_t
\end{aligned}$$

Primero simplificamos el coeficiente de x_t:

$$\begin{aligned}
\text{Coef}_{x_t} &= \frac{1}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} \left[\sqrt{\alpha_t}(1 - \tilde{\alpha}_{t-1}) \sqrt{\tilde{\alpha}_t} + \sqrt{\tilde{\alpha}_{t-1}} \beta_t \right] \\
&= \frac{1}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} \left[\alpha_t \sqrt{\tilde{\alpha}_{t-1}}(1 - \tilde{\alpha}_{t-1}) + \sqrt{\tilde{\alpha}_{t-1}} \beta_t \right] && (\because \tilde{\alpha}_t = \alpha_t \tilde{\alpha}_{t-1}) \\
&= \frac{\sqrt{\tilde{\alpha}_{t-1}}}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} [\alpha_t - \alpha_t \tilde{\alpha}_{t-1} + \beta_t] \\
&= \frac{\sqrt{\tilde{\alpha}_{t-1}}}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} [1 - \alpha_t \tilde{\alpha}_{t-1}] && (\because \alpha_t + \beta_t = 1) \\
&= \frac{\sqrt{\tilde{\alpha}_{t-1}}}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} [1 - \tilde{\alpha}_t] \\
&= \frac{\sqrt{\tilde{\alpha}_{t-1}}}{\sqrt{\tilde{\alpha}_t}} = \frac{\sqrt{\tilde{\alpha}_{t-1}}}{\sqrt{\alpha_t \tilde{\alpha}_{t-1}}} = \frac{1}{\sqrt{\alpha_t}}
\end{aligned}$$

Ahora simplificamos el coeficiente de ϵ_t:

$$\text{Coef}_{\epsilon_t} = -\frac{\sqrt{\tilde{\alpha}_{t-1}} \beta_t \sqrt{1 - \tilde{\alpha}_t}}{(1 - \tilde{\alpha}_t)\sqrt{\tilde{\alpha}_t}} = -\frac{\beta_t \sqrt{\tilde{\alpha}_{t-1}}}{\sqrt{1 - \tilde{\alpha}_t}\sqrt{\alpha_t \tilde{\alpha}_{t-1}}} = -\frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1 - \tilde{\alpha}_t}}$$

Sustituyendo los coeficientes simplificados de nuevo en $\mu(x_t, x_0)$:

$$\mu(x_t) = \frac{1}{\sqrt{\alpha_t}} x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1 - \tilde{\alpha}_t}} \epsilon_t$$

Factorizando $\frac{1}{\sqrt{\alpha_t}}$:

$$\mu(x_t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \tilde{\alpha}_t}} \epsilon_t \right) \quad (6)$$

.

Ejercicio 3.

Usando una imagen de referencia de un cóndor, Fig. 1, se implemento el proceso directo con la formula que obtuvimos haciendo la expansión recursiva en el ejercicio 1:

$$x_t = \sqrt{\tilde{\alpha}_t}x_0 + \sqrt{1-\tilde{\alpha}_t}\epsilon_t$$

Figura 1: Imagen de referencia usada.

El proceso directo se implemento para un valor fijo que cambia linealmente desde $\beta_0 = 0{,}0001$ hasta $\beta_T = 0{,}02$ con $T = 200$. En la Fig. 2 se puede observar la degradación de la imagen a medida que se la agrega ruido para $t \in [0, 3, 10, 30, 100, 200]$.

Figura 2: Imagen de referencia con ruido.

Ejercicio 4, 5 y 6.

Se implemento una red neuronal con arquitectura Unet incluyendo el tiempo embebido como entrada adicional. Para esto se definió una clase *TimeEmbedding(nn.Module)*.

El encoder posicional implementado es el utilizado en el *paper*: *Attention is All You Need*. El vector de codificación posicional PE_t para un paso de tiempo t y una dimensión i se calcula mediante las siguientes ecuaciones:

$$\mathrm{PE}_{(t,2i)} = \sin\left(\frac{t}{10000^{i/(halfdim-1)}}\right)$$

$$\mathrm{PE}_{(t,2i+1)} = \cos\left(\frac{t}{10000^{i/(halfdim-1)}}\right)$$

Donde:

- t: Es el paso de tiempo de difusión actual $(1 \leq t \leq T)$.
- i: Es el índice de la dimensión dentro del vector de codificación $(0, 1, 2, \ldots, halfdim-1)$.
- $halfdim$: Es la mitad de la dimensión de salida del embedding de tiempo (el tamaño final del vector).

Este vector sinusoidal PE_t es posteriormente transformado mediante una *Red Neuronal Multicapa (MLP)* antes de ser inyectado en las capas del U-Net.

Luego, se definió la función de perdida, muestreando el intervalo de tiempos [0,T] de manera uniforme, como:

$$L = ||\epsilon - \epsilon_\theta(x_t(x_0, \epsilon), t)||^2$$

donde $\epsilon \sim \mathcal{N}(0, 1)$ y $x_t(x_0, \epsilon)$ esta generado por el proceso directo.

Con esto, se entreno la red usando los datos de MNIST y se evaluó su performance en términos de la capacidad de eliminar el ruido de una muestra que paso por el proceso directo. El numero total de pasos usado para el proceso de difusión fue $T = 200$, con β_t aumentando linealmente desde $\beta_0 = 0{,}0001$ hasta $\beta_T = 0{,}02$. El optimizador utilizado fue *Adam* con *learning rate = 1e-4*. En la red Unet se usaron *in_channels=1* (canales de la imagen, 1 por ser escala de grises), *base_channels=64* (numero de canales iniciales de la Unet), y *time_emb_dim=256* (dimensión del embedding sinusoidal de tiempo).

En la Fig. 3 se puede observar la evolución de la loss durante el entrenamiento. Se entreno hasta 200 épocas. El numero de épocas se eligió porque demoraba mucho el entrenamiento y al realizar la reconstrucción de los dígitos a partir de ruido daba resultados razonables.

A continuación, se evaluó la capacidad de eliminar el ruido de una imagen de muestra. Para esto se tomo una imagen del dataset del MNIST, y con el proceso directo se le aplico ruido hasta un cierto t elegido. Luego, la imagen x_t con ruido se la uso en el modelo entrenado para predecir ϵ_θ. Recordando la ecuación del proceso directo que relaciona x_t y x_0 con el ruido agregado ϵ_t:

$$x_t = \sqrt{\tilde{\alpha}_t} x_0 + \sqrt{1 - \tilde{\alpha}_t} \epsilon_t$$

Despejando x_0 tenemos:

$$x_0 = \frac{1}{\sqrt{\tilde{\alpha}_t}} \left(x_t - \sqrt{1 - \tilde{\alpha}_t} \epsilon_t \right)$$

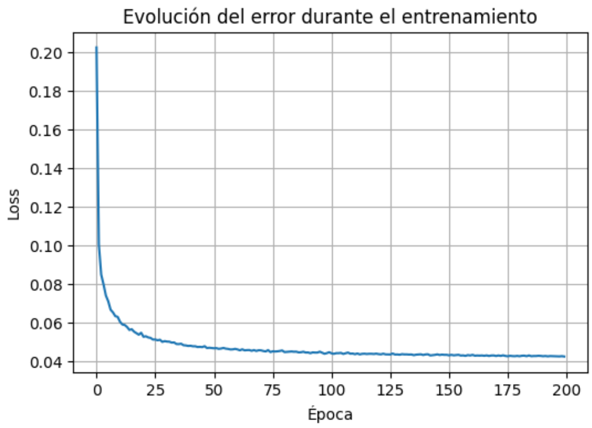
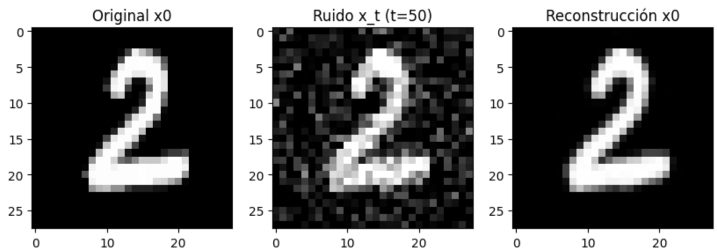
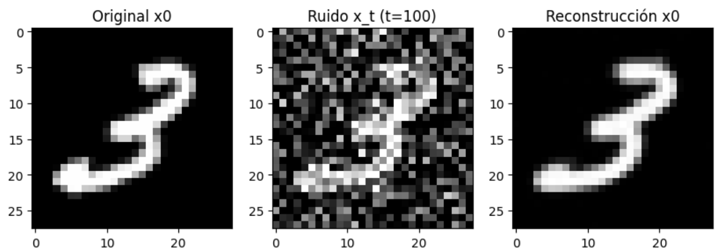


Figura 3: Loss en función de las épocas.

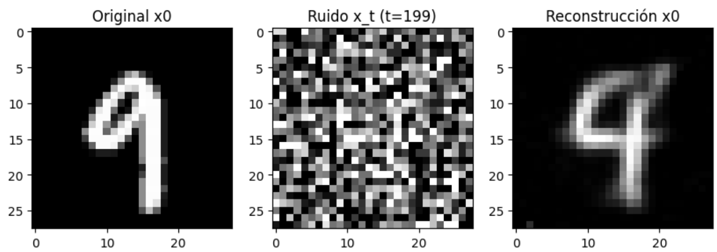
Usando esta ultima expresión con el ruido ϵ_θ predicho por el modelo, se reconstruyo la imagen original. En la Fig. 4 se pueden observar los resultados obtenidos de este proceso. Se observa que para un ruido alto ($t = 199$) la reconstrucción no es media difusa, aunque preserva las características principales del dígito. Para un ruido intermedio ($t = 100$), la reconstrucción es casi idéntica a la imagen original, y con un ruido bajo ($t = 50$), también. Se puede ver que la red aprendió a predecir el ruido de las imágenes.



(a) Ejemplo de reconstrucción con ruido $t = 50$.



(b) Ejemplo de reconstrucción con ruido $t = 100$.



(c) Ejemplo de reconstrucción con ruido $t = 199$.

Figura 4: Ejemplos de reconstrucción con diferente nivel de ruido.

Finalmente, se probo generar nuevos dígitos de MNIST a partir de ruido. Para esto se partió desde una imagen generada de ruido $\mathcal{N}(\mathbf{0}, \mathbf{I})$, y se aplico el proceso inverso:

$$x_{t-1} = \mu(x_t) + \sqrt{\beta_t}\epsilon_t$$

donde

$$\mu(x_t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\tilde{\alpha}_t}}\epsilon_t\right)$$

.

De esta forma, partiendo de $t = 200$ hasta $t = 0$, se fue calculando ϵ_θ predicho por el modelo para la dada imagen x_t. Calculando luego con esto el $x_(t-1)$ siguiente, y así sucesivamente. En la Fig. 5 se puede observar el proceso para diferentes imágenes de ruido generadas. Se ve que algunas imágenes logran recrear bastante bien los dígitos del MNIST, mientras que otras presentan trazos que asemejan números pero no son completamente correctos.

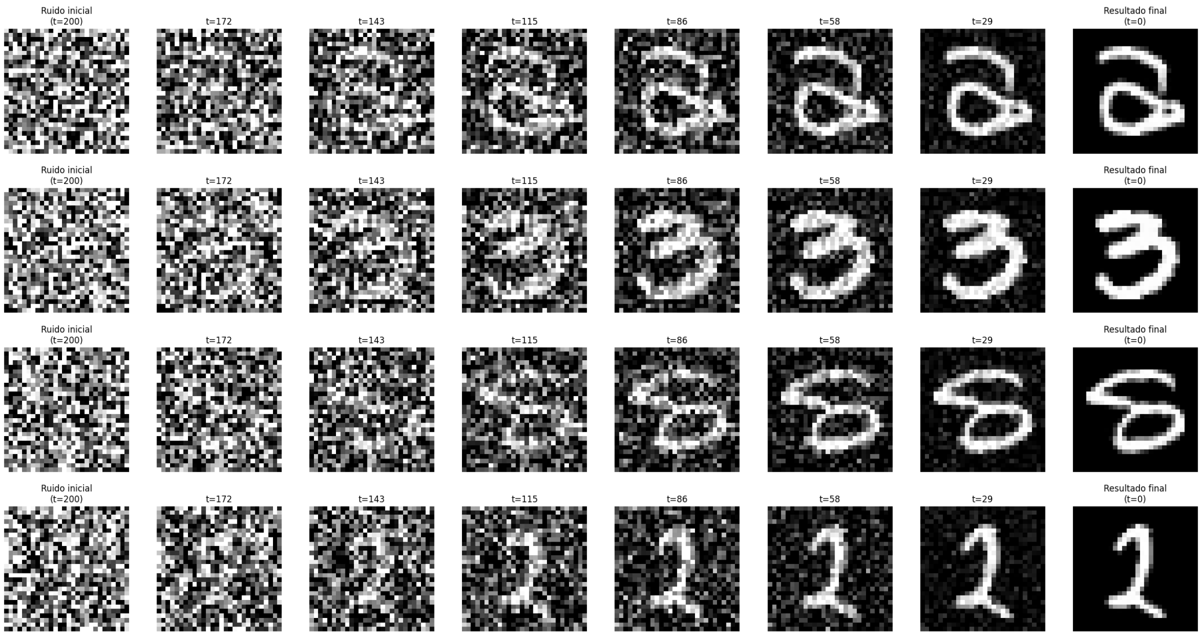


Figura 5: Ejemplos de creación de dígitos de MNIST partiendo desde ruido.

2. Apéndice

Los códigos con los ejercicios resueltos se encuentran en el repositorio: `https://github.com/AnaMorresi/DeepLearning`.