DEEP LEARNING PRÁCTICA 5 - 2025

Modelos de Difusión

1. Proceso *directo*: considerar un proceso iterativo con condición inicial $\mathbf{x}_0$ y que en cada paso de tiempo evoluciona de acuerdo a

$$\mathbf{x}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta}_t\epsilon_t \tag{1}$$

   donde ϵ_t es una variable gaussiana con media 0 y varianza 1 independiente para cada tiempo y β_t son coneficientes positivos y menores que 1.

   - ¿Cual de la distribución de probabilidad de $\mathbf{x}_t$ para $t > 0$?
   - Demostrar que su media es $\sqrt{\tilde{\alpha}_t}\mathbf{x}_0$ y su matriz de covarianza es $(1-\tilde{\alpha}_t)I$ para todo t donde $\tilde{\alpha}_t = \prod_{t' \leq t} \alpha_{t'}$ y $\alpha_t = 1 - \beta_t$

2. Proceso *inverso*: definimos el proceso inverso por la relación de recurrencia

$$\mathbf{x}_{t-1} = \mu(\mathbf{x}_t) + \sqrt{\beta_t}\epsilon_t \tag{2}$$

   donde la función $\mu(\mathbf{x})$ denota la media del proceso inverso.

   Este proceso tambien se puede pensar condicionado a $\mathbf{x}_0$:

$$\mathbf{x}_{t-1} = \mu(\mathbf{x}_t, \mathbf{x}_0) + \sqrt{\beta_t}\epsilon_t \tag{3}$$

   Esta relación define $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$

   - Utilizando la regla de Bayes calcular esta cantidad en término de $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)$, $p(\mathbf{x}_t|\mathbf{x}_0)$ y $p(\mathbf{x}_{t-1}|\mathbf{x}_0)$
   - Usando las probabilidades del proceso directo para estas tres probabilidades demostrar que la media del proceso inverso condicionado es

$$\mu(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\tilde{\alpha}_{t-1})}{1-\tilde{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\tilde{\alpha}_{t-1}}\beta_t}{1-\tilde{\alpha}_t}\mathbf{x}_0 \tag{4}$$

   - Usando que el proceso directo se puede escribir como

$$\mathbf{x}_t = \sqrt{\tilde{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\tilde{\alpha}_t}\epsilon_t \tag{5}$$

eliminar $\mathbf{x}_0$ y probar que la media del proceso inverso se puede escribir como

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

- Este resultado motiva la propuesta de

$$\mu_\theta(\mathbf{x}_t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \tilde{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t) \right) \tag{7}$$

  donde $\epsilon_\theta(\mathbf{x}_t)$ es una función implementada con una rad neuronal para "aprender" la variable aleatoria necesaria para invertir la inyección de ruido. Observar que el sistema requiere conocer el tiempo t, asi que si queremos usar una sola red neuronal debemos embeber esa variable en la entrada de la red: $\epsilon_\theta(\mathbf{x}_t) \rightarrow \epsilon_\theta(\mathbf{x}, t)$

3. Bajar la imagen de referencia del condor del repositorio e implementar el prooceso directo con un valor fijo de que ccambia linealmente en tiempo desde $\beta_0 = 0.0001$ hasta $\beta_T = 0.02$, con $T = 200$. Examinar la degradación a medida que se agraga ruido. Graficar $\mathbf{x}_0$, $\mathbf{x}_3$, $\mathbf{x}_{10}$, $\mathbf{x}_{30}$, $\mathbf{x}_{100}$, $\mathbf{x}_{200}$.

4. Programar una red neuronal con arquitectura Unet que incluya el tiempo embebido como entrada adicional. Definir una clase TimeEmbedding(nn.Module).

5. Escribir una función de pérdida muestreando el intervalo de tiempos $[0, T]$ de manera uniforme y definiendo

$$L = ||\epsilon - \epsilon_\theta(\mathbf{x}_t(\mathbf{x}_0, \epsilon), t)||^2 \tag{8}$$

   donde ϵ es una variable gaussiana con media 0 y varianza 1 y $\mathbf{x}_t(\mathbf{x}_0, \epsilon)$ está generado por el proceso directo.

6. Entrenar la red usando los datos de mnist y evaluar su performance en términos de la capacidad de eliminar el ruido de una muestra que pasó por le proceso directo.