



# Trabajo Práctico Final

Entregable Autoencoders Variacionales y Aprendizaje Semi-Supervisado

1 de Agosto del 2025

Modelos Generativos Profundos para Imágenes - ECI M3

Integrante	LU	Correo electrónico
Ranieri, Martina	1118/22	martubranieri@gmail.com



## Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja) Intendente Güiraldes 2610 - C1428EGA Ciudad Autónoma de Buenos Aires - Rep. Argentina Tel/Fax: (++54+11) 4576-3300

http://www.exactas.uba.ar

# Índice

1.	Problema 1: Implementación del Autoencoder Variacional (VAE)	2
	1.1. Pregunta 3: Reporte media y el error estándar	
	1.3. Pregunta 5: Variante del VAE, $\beta$ -VAE	
2.	Problema 2: Implementación del VAE de Mezcla de Gaussianas (GMVAE)	3
	2.1. Pregunta 3: Reporte media y el error estándar	
3.	Problema 3: Implementación del VAE Semi-Supervisado (SSVAE)	4
	3.1. Precisión de clasificación: supervisado vs semi-supervisado	4
4.	Conclusión	4

### 1. Problema 1: Implementación del Autoencoder Variacional (VAE)

### 1.1. Pregunta 3: Reporte media y el error estándar

	Media	Error estándar
NELBO	101.31	0.08
Regularización	19.33	0.04
Reconstrucción	81.98	0.10

Tabla 1

Respuesta. Se realizaron 10 ejecuciones y los resultados de media y desvío estándar son los mostrados en la Tabla 1. Esto sugiere que el modelo tiene un comportamiento estable con distintas inicializaciones ya que presenta un valor bajo del error estándar. La ELBO negativa se encuentra cercana al valor esperado (alrededor de 100), lo que indica que el modelo logró un compromiso razonable entre la reconstrucción de los datos y la regularización del espacio latente.

### 1.2. Pregunta 4: Visualizar 200 dígitos muestreados de $p_{\theta}(x)$

**Respuesta.** Se generaron 200 imágenes muestreadas desde el prior latente p(z) del modelo en la Figura 1. Este proceso consiste en:

- 1. Muestrear vectores latentes  $z \sim p(z)$ , sin usar datos reales.
- 2. Generar imágenes  $x \sim p_{\theta}(x|z)$  pasando esos vectores por el decoder del modelo.
- 3. Las imágenes generadas representan muestras producidas por el modelo entrenado.

Con esto estamos probando si el modelo aprendió bien la distribución de los datos y usar lo aprendido para generar imágenes nuevas sin copiar ni reconstruir nada del dataset.

En la Figura 1 podemos observar que el modelo es capaz de generar muestras coherentes con el dominio de los datos originales, aunque algunas imágenes todavía presentan ambigüedad o ruido. Esto puede ocurrir ya que el VAE realiza un compromiso entre calidad de reconstrucción y regularización: al exigir que las distribuciones latentes se parezcan al prior, se sacrifica parte de la precisión en la reconstrucción, lo que a veces genera muestras menos definidas.

vae generated samples from the set prior

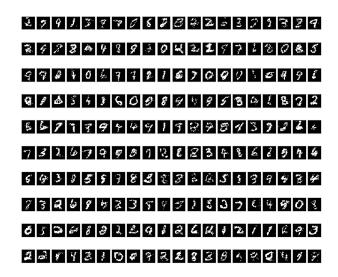


Figura 1: El prior p(z) es una normal estándar:  $\mathcal{N}(0, I)$ .

#### 1.3. Pregunta 5: Variante del VAE, $\beta$ -VAE

Dar una interpretación intuitiva del impacto de  $\beta$  en la optimización del VAE; específicamente, ¿qué sucede cuando  $\beta$  = 1? ¿Qué pasa cuando  $\beta$  se aumenta por encima de 1? Para esta pregunta, será suficiente una descripción cualitativa.

$$\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log p_{\theta}(\mathbf{x}|\mathbf{z}) \right] - \beta D_{\text{KL}} \left( q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}) \right) \tag{1}$$

Figura 2: Función objetivo de  $\beta$ -VAE.  $\beta$  es un real positivo

**Respuesta.** Primero, recordemos que el primer término corresponde al de reconstrucción, el cual mide qué tan parecidas son las imágenes reconstruidas a las originales. Cuanto más bajo sea, tenemos menos pérdida de información y por ende una buena calidad de reconstrucción. Luego, el segundo término mide cuánto difieren las distribuciones latentes que produce el encoder  $q_{\phi}(z|x)$  del prior p(z). Esta regularización ayuda a que el espacio latente sea ordenado y generalizable.

En esta variante  $\beta$ -VAE, cuando  $\beta = 1$  tenemos la misma ecuación original de VAE que vimos, manteniendo un balance entre ambos términos. Ahora, cuando  $\beta > 1$  se le da mayor peso al término de regularización y fuerza a que la distribuciones

latentes  $q_{\phi}(z|x)$  se parezcan más al prior p(z) ya que se le da mayor importancia a este término en la minimización. Entonces, el modelo cede en reconstrucción para acercarse más al prior y esto puede significar imágenes más borrosas o con menos detalles.

# 2. Problema 2: Implementación del VAE de Mezcla de Gaussianas (GM-VAE)

### 2.1. Pregunta 3: Reporte media y el error estándar

	Media	Error estándar
NELBO	99.06	0.15
Regularización	17.18	0.06
Reconstrucción	81.25	0.15

Tabla 2

Respuesta. Se realizaron 10 ejecuciones y los resultados de media y desvío estándar son los mostrados en la Tabla 2. Podemos observar de que la función de pérdida total (NELBO) es menor a la que teníamos con VAE, esto posiblemente suceda ya que el usar mezcla de Gaussianas nos permite capturar mejor las distribuciones latentes. Por otro lado, el término de regulación (KL) también es menor al valor obtenido en VAE. Estos resultados nos sugieren que la eficiencia del codificador es mejor, y el tener un valor de regulación similar, nos indica que el GMVAE logra una representación latente más expresiva sin comprometer significativamente la calidad de las reconstrucciones.

### 2.2. Visualizar 200 dígitos muestreados de $p_{\theta}(x)$

Respuesta. Se generaron 200 imágenes muestreadas desde el prior latente p(z) del modelo en la Figura 3 al igual que en la Figura 1. Para comparar ambos modelos, se utilizó un clasificador provisto en los archivos del TP para el dataset MNIST. Fue entrenado sobre el conjunto de entrenamiento, y al evaluarlo con el conjunto de prueba obtuvo un accuracy de 97.66%, es decir que es efectivo para reconocer elementos del dataset.

Este clasificador se utilizó para asignar etiquetas (0–9) a las muestras generadas por el VAE y el GMVAE, contabilizando cuántas veces se asignó cada una. Esto mide la diversidad de dígitos generados. Si el modelo solo genera imágenes que el clasificador interpreta como '6' y '8', va a tener un sesgo. Para que un modelo sea bueno, se espera que los conteos estén distribuidos de forma pareja, idealmente cercanos a 20 ya que se generaron 200 imágenes (200/10), y que no haya dígitos ausentes o muy poco frecuentes.

Dígito generado	Cantidad VAE	Cantidad GMVAE
0	12	26
1	12	15
2	26	22
3	30	25
4	23	17
5	15	17
6	17	20
7	19	14
8	20	19
9	26	25

El conteo de los dígitos muestra que el GMVAE tiende a producir una distribución más equilibrada entre las clases, mientras que el VAE presenta mayor sesgo hacia ciertos dígitos (por ejemplo, el '3'). Esto podría indicar que el prior del GMVAE permite cubrir mejor el espacio latente, generando imágenes de más clases, aunque no necesariamente implica mayor calidad visual en cada dígito.

gmvae generated samples from the set prior

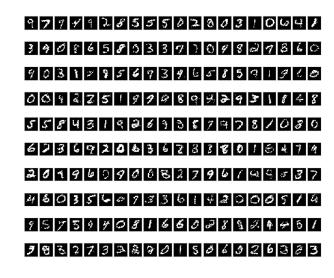


Figura 3: El prior p(z) es una mezcla de gaussianas

## 3. Problema 3: Implementación del VAE Semi-Supervisado (SSVAE)

### 3.1. Precisión de clasificación: supervisado vs semi-supervisado

	Supervisado	Semi-supervisado
Accuracy	75.97%	95.45%

Tabla 3: Resultados de la ejecución con 30000 iteraciones

Respuesta. En la Tabla 3 vemos que al entrenar únicamente en el escenario supervisado (ignorando los datos no etiquetados) se obtuvo una precisión de clasificación del 75.97 %, mientras que al incluir el término no supervisado con SSVAE, la precisión mejoró hasta llegar al 95.45 %. Esto confirma que el uso de datos no etiquetados aporta información valiosa y mejora notoriamente el desempeño del modelo ya que permite aprender representaciones latentes más expresivas y aprovechar mejor la estructura del conjunto de datos, incluso cuando la cantidad de etiquetas disponibles es reducida.

### 4. Conclusión

A lo largo de este trabajo se implementaron y analizaron diferentes variantes de autoencoders variacionales: el VAE estándar, el GMVAE y el SSVAE. En primer lugar, el VAE mostró un equilibrio entre los términos de reconstrucción y regularización, siendo capaz de generar muestras nuevas pero con cierta pérdida de nitidez o claridad en las imágenes. Luego, el GMVAE permitió una representación latente mejor y diversa, reduciendo el sesgo en la generación de clases y capturando mejor la estructura multimodal de los datos. Finalmente, el SSVAE mostró la ventaja de incorporar datos no etiquetados en un esquema semi-supervisado, logrando una mejora sustancial en la precisión de clasificación respecto al escenario completamente supervisado.

Estos resultados nos muestran el valor de los modelos generativos profundos como herramientas para aprender representaciones latentes útiles en tareas tanto de generación como de clasificación. En particular, el caso del SSVAE muestra que el uso eficiente de grandes cantidades de datos no etiquetados puede compensar la escasez de etiquetas, lo cual resulta especialmente relevante en aplicaciones reales donde obtener las etiquetas suele ser costoso o impráctico.