

Autoencoder Variacional con Covarianza Factorizada para Imágenes Astronómicas

Javier Rojas Herrera

Universidad Austral de Chile, General Lagos 2086, Valdivia, Chile

Resumen En 2022 comenzará a operar el Legacy Survey of Space and Time (LSST), sondeo astronómico de gran envergadura que producirá decenas de TeraBytes de datos cada noche. LSST permitirá estudiar en gran detalle los objetos transientes, eventos astronómicos que ocurren en distintas escalas y luego se desvanecen, como las Supernovae o los Gamma Ray Bursts. Métodos de aprendizaje de máquinas no supervisados que operen sobre imágenes de diferencia podrían permitir descubrir nuevos tipos de objetos transientes en el stream de datos nocturno de LSST. En este trabajo se diseña e implementa una red neuronal no supervisada de tipo autoencoder variacional con capas convolucionales para aprender representaciones latentes de imágenes astronómicas de diferencia. La innovación radica en la incorporación de una aproximación de la covarianza de los datos en la red neuronal, lo cual permite modelar correlaciones entre píxeles vecinos de forma directa. Esto repercute en la calidad de las imágenes generadas por el modelo y en el espacio latente aprendido. Los resultados preliminares muestran que el modelo de covarianza completa alcanza un menor error de reconstrucción y genera imágenes más realistas que los métodos convencionales al ser entrenado en un dataset de imágenes cotidianas. Los siguientes pasos son entrenar el modelo en datos astronómicos obtenidos del proyecto HiTS y mejorar la estabilidad numérica del modelo a través de regularización.

Keywords: Redes Neuronales Artificiales · Autoencoder Variacional · Astronomía.

1. Introducción

Los avances tecnológicos recientes han permitido la construcción de observatorios astronómicos de gran escala, muchos de estos instalados en el norte grande de Chile. Estos observatorios sinópticos se caracterizan por observar el cielo completo cada noche con el objetivo de estudiar los objetos que cambian en el tiempo. Un ejemplo emblemático es el Vera Rubin Observatory (VRO) que en 2022 comenzará su campaña de colección de datos denominada Legacy Survey of Space and Time (LSST) [11]. LSST cubrirá un área de 18.000 grados cuadrados equivalente a todo el cielo nocturno del hemisferio sur. Cada noche LSST generará 20 Terabytes de imágenes astronómicas que luego de sus 10 años de operación resultarán en un catálogo de 550 PetaBytes representando 37 billones

de objetos astronómicos [11]. En este escenario cualquier tipo de procesamiento o análisis manual por parte de astrónomos es infactible. La necesidad por algoritmos computacionales para el análisis eficiente y oportuno de estos datos llama a la colaboración entre astrónomos e ingenieros informáticos.

Un problema astronómico de suma relevancia hoy en día es la identificación rápida de fenómenos transientes, siendo un ejemplo emblemático de estos las Supernovae (SNe), que corresponden a la explosión de una gigante roja al terminar su ciclo de fusión. La detección de estos eventos es vital para avanzar la investigación en astrofísica y cosmología. Se espera que gracias a las SNe detectadas utilizando datos como los recolectados por LSST se podrán confirmar o refutar las múltiples teorías actuales sobre la naturaleza de energía oscura [19].

Las SNe fueron el objeto principal de estudio del sondeo High Cadence Transient Survey (HiTS) [8], realizado entre 2013 y 2015 utilizando el instrumento Dark Energy Camera (DECam) instalado en Cerro Tololo la cuarta región de Chile. Este sondeo se puede considerar como un precursor del LSST, siendo de similares características, pero en una escala 100 veces menor. Los datos generados por HiTS corresponden a cientos de miles de secuencias de imágenes en torno a objetos candidatos a transiente estelar, entre los que se encuentran supernovae, pero también estrellas variables, asteroides y mayoritariamente artefactos ocasionados por errores del telescopio o del programa de reducción de imágenes.

Para clasificar los datos de HiTS se realizó un arduo proceso de simulación de datos y *feature engineering* con el fin de obtener etiquetas confiables y caracterizar adecuadamente los datos [8], respectivamente. Los features o características diseñadas se basaron en técnicas de procesamiento de imágenes clásico y estadística. Una posible dificultad asociada al diseño de features es que podrían no ser directamente transferibles para aplicarse en otros proyectos de sondeo. Otra desventaja es que el cálculo de características no escala en términos de costo computacional para flujos de datos de mayor velocidad y volumen.

Hoy en día existen modelos supervisados para la clasificación de transientes astronómicos a partir de imágenes [3]. La principal dificultad en este caso es la necesidad de un gran conjunto de datos etiquetados. El proceso de etiquetado puede ser altamente costoso en términos de tiempo y dinero. Por otro lado, la aumentación de datos sintéticos podría introducir sesgos importantes en el dataset. Esto nos motiva a utilizar métodos no supervisados que no requieran etiquetas para extraer la información fundamental de los datos.

En particular, detectamos una oportunidad de aportar desde la disciplina informática en la generación automática de características de baja dimensión para imágenes astronómicas que sean informativas, interpretables y escalen computacionalmente, adelantándonos de esta forma al advenimiento de los datos del LSST. Para esto proponemos adaptar redes neuronales profundas no supervisada de tipo autoencoder variacional [14] que serán entrenadas sobre las imágenes de diferencia de HiTS. Proponemos extender el modelo clásico incorporando una versión factorizada de la covarianza completa con el objetivo de tomar en con-

sideración la correlación existente entre los píxeles y en consecuencia mejorar el modelamiento de las imágenes astronómicas.

1.1. Pregunta de investigación e Hipótesis

En este trabajo se intenta responder las siguientes interrogantes: ¿Modelar la covarianza completa factorizada en el VAE mejora las imágenes generadas? ¿Es el espacio latente más representativo al mejorar las reconstrucciones del VAE? ¿Es posible replicar el modelo propuesto en diferentes áreas que trabajen con imágenes?

La hipótesis principal de este trabajo es que modelar la covarianza del decodificador del autoencoder variacional por medio de factorizaciones de bajo rango permitirá extraer características a partir de imágenes astronómicas con las que se alcanzará un menor error de reconstrucción y clasificación en comparación al modelo de autoencoder variacional clásico.

2. Estado del arte

El autoencoder [1,9] es una arquitectura de red neuronal artificial no-supervisada que consiste de un par codificador $g_\phi(\cdot)$ y decodificador $f_\theta(\cdot)$, con parámetros ϕ y θ , respectivamente. El autoencoder se entrena típicamente con el objetivo de minimizar el error cuadrático medio entre la entrada $x \in \mathbb{R}^D$ y la salida

$$\min L(\theta, \phi) = \sum_{i=1}^N \|x_i - f_\theta(g_\phi(x_i))\|^2, \quad (1)$$

donde $z = g_\phi(x) \in \mathbb{R}^K$ con $K \ll D$ es el cuello de botella o código del autoencoder. La minimización del error de reconstrucción fuerza a que el código guarde la información más importante de la entrada, *i.e.* aquella mínima información que se requiere para reconstruirla. El autoencoder es un modelo ampliamente usado como extractor de características en aplicaciones con imágenes [15,20].

En la práctica otros costos como la entropía cruzada o la perceptual loss [12] pueden considerarse dependiendo del tipo de dato que se utilice. La optimización de la función de costo se realiza frecuentemente usando métodos de primer orden basados en gradiente descendente estocástico con tasa de aprendizaje adaptativa [13]. En particular para el caso de las imágenes, es conveniente utilizar capas convolucionales en las redes codificadora y decodificadora [17]. La única salvedad es que se debe utilizar la operación de convolución traspuesta en el decodificador con tal de aumentar la dimensionalidad [16].

Uno de los avances más reciente en torno a las arquitecturas de tipo autoencoder es el Autoencoder Variacional o *Variational Autoencoder* (VAE) [14,21]. A diferencia del autoencoder convencional, el código del VAE es una variable estocástica que se asume una Gaussiana con covarianza diagonal

$$z \sim q_\phi(z|x) = \mathcal{N}(z \mid \mu_\phi(x), \sigma_\phi^2(x)), \quad (2)$$

donde $\mu_\phi(x)$ y $\sigma_\phi^2(x)$ son redes neuronales profundas. Este supuesto de independencia de las variables latentes es propio de los algoritmos de inferencia variacional de tipo *mean-field* [2]. Para el decodificador se asume usualmente una distribución Gaussiana con covarianza esférica

$$x \sim p_\theta(x|z) = \mathcal{N}(x \mid \mu_\theta(z), \hat{\sigma}^2 I), \quad (3)$$

donde $\mu_\theta(z)$ es una red neuronal profunda y $\hat{\sigma}^2$ es un hiper-parámetro escogido por el usuario. El modelo se entrena usando inferencia variacional estocástica [10], resultando en la siguiente función de costo

$$\text{máx } \mathcal{L}(\theta, \phi) = \sum_{i=1}^N \mathbb{E}_{z \sim q_\phi(z|x_i)} [\log p_\theta(x_i|z)] - D_{KL}[q_\phi(z|x_i) \| p(z)], \quad (4)$$

que teóricamente corresponde a una cota superior para la evidencia del modelo $\mathbb{E}[\log p(x)]$, obtenida de aproximar el posterior inferencial $p(z|x)$ por el posterior simplificado $q(z|x)$ [14].

Típicamente se supone un prior Gaussiano estándar para la variable latente $p(z) = \mathcal{N}(0, I)$ con lo cual el término de la derecha de la ecuación (4) tiene forma analítica cerrada

$$D_{KL}[q_\phi(z|x_i) \| p(z)] = -\frac{1}{2} (1 + \log \sigma_\phi(x_i)^2 - \sigma_\phi(x_i)^2 - \mu_\phi(x_i)^2). \quad (5)$$

La minimización de (5) busca que el posterior aproximado de la distribución inferencial se asemeje al prior, actuando como un regularizador para el modelo.

El término de la izquierda de la Ecuación (4) se calcula mediante estimación de Monte-Carlo empleando el truco de reparametrización [14]

$$z = \mu_\phi(x) + \varepsilon \sigma_\phi(x), \quad \varepsilon \sim \mathcal{N}(0, I) \quad (6)$$

resultando en una expresión que es derivable en función de θ y ϕ

$$\begin{aligned} \mathbb{E}_{z \sim q_\phi(z|x)} [\log p_\theta(x|z)] &= \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, I)} [\log p_\theta(x \mid \mu_\phi(x) + \varepsilon \sigma_\phi(x))] \\ &\approx -\frac{D}{2} \log(2\pi \hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^J \|x - \mu_\theta(\mu_\phi(x) + \varepsilon_j \sigma_\phi(x))\|^2, \end{aligned} \quad (7)$$

donde se muestrea J veces de la variable ε . Notamos que maximizar la Ecuación (7) para el caso $\hat{\sigma} = 1$ es equivalente a minimizar el error cuadrático medio.

Las ventajas de VAE sobre el autoencoder tradicional es que al ser un modelo generativo es posible muestrear desde el espacio latente estocástico y generar datos sintéticos nuevos. Adicionalmente al incorporar un prior en el código, el VAE se regulariza de forma natural obteniendo un espacio latente con mayor interpretabilidad. La limitación del VAE yace en los supuestos que impone sobre la distribución de inferencial y generativa.

En [6,7] se observó que el error entre la entrada y la reconstrucción de un VAE entrenado sobre imágenes naturales presenta estructura, es decir que no se

cumple supuesto de independencia. En general las imágenes tienen alta correlación espacial entre sus píxeles [4], y para considerar esto [6] propone utilizar una distribución Gaussiana con covarianza completa para el decoder. Sin embargo, modelar la covarianza completa tiene dos dificultades importantes: (1) no es escalable en el sentido de la cantidad de parámetros que se incorporan y (2) dificulta el proceso de optimización debido a las restricciones que se incorporan.

En la práctica se utiliza una factorización de bajo rango de la matriz de covarianza y factores de regularización adicionales [6], obteniendo mejores resultados en generación de imágenes naturales con respecto al VAE convencional [14]. En la sección siguiente se describe en detalle el modelo de covarianza factorizada que se utiliza en este trabajo para el modelamiento de imágenes astronómicas.

3. Metodología

3.1. Modelo

En este trabajo se propone utilizar una distribución generativa Gaussiana con covarianza completa

$$p_{\theta}(x|z) = N(x \mid \mu_{\theta}(z), \Sigma_{\theta}(z)) \quad (8)$$

donde μ es la media y Σ es la covarianza. Esto difiere del supuesto tradicional de VAE donde los elementos de x son independientes (covarianza identidad) [14]. En el caso de las imágenes un modelo que asume independencia podría ser adecuado para modelar píxeles muy separados, pero no así para píxeles vecinos que se caracterizan por tener una fuerte correlación [4]. La dificultad radica en cómo definir dicha vecindad. Una vecindad pequeña resulta en un modelo poco realista mientras que una vecindad grande resulta en una covarianza densa muy difícil de estimar directamente.

Una alternativa es trabajar con la matriz de precisión $\Lambda = \Sigma^{-1}$, es decir el inverso de la covarianza. Un cero en la matriz de covarianza equivale a asumir independencia entre dos elementos, mientras que un cero en la matriz de precisión equivale a asumir independencia condicional de dos elementos dados los demás [5]. En este trabajo asumiremos que si dos píxeles no son inmediatamente vecinos entonces son condicionalmente independientes resultando en una matriz de precisión *sparse* que es fácil de manipular y que se traduce en una covarianza densa [6]. Otra ventaja de trabajar con la matriz de precisión es que no es necesario invertir Σ para estimar el logaritmo de la verosimilitud, lo cual disminuye considerablemente el costo computacional.

Se considera una descomposición de Cholesky de la matriz de precisión

$$\Lambda = LL^T, \quad (9)$$

donde L es una matriz triangular inferior, con lo cual basta forzar a los elementos de la diagonal de L a ser no-negativos para asegurar que la matriz de

covarianza es semi-definida positiva. Con esto podemos escribir el logaritmo de la verosimilitud como

$$\log p_\theta(x|z) = -\frac{D}{2} \log(2\pi) + \sum_{i=1}^d \log L_\theta[i, i] - \frac{1}{2} \|L_\theta(z)(x - \mu_\theta(z))\|^2 \quad (10)$$

donde en este caso $L_\theta(z)$ y $\mu_\theta(z)$ son redes neuronales profundas. La función de costo final del modelo propuesto se obtiene de reemplazar las ecuaciones (10) y (5) en la Ecuación (4). El modelo resultante es una generalización del VAE [14] puesto que

- Si $L_\theta(z) = \hat{\sigma}I$, con $\hat{\sigma}$ escalar, entonces se recupera el VAE convencional con covarianza esférica que no modela dependencias entre píxeles.
- Si $L_\theta(z) = \hat{\sigma}(z)I$, con $\hat{\sigma} \in \mathbb{R}_+^D$, entonces se tiene un VAE que no modela dependencias entre píxeles, pero que permite que la escala de cada píxel varíe libremente.

En la siguiente sección se describen los datos que se usan para evaluar el modelo con covarianza completa y compararlo con los modelos simplificados.

Con respecto a la generación a partir del decodificador notar que se cumple

$$\Sigma = \Lambda^{-1} = L^{-1^T} L^{-1}, \quad (11)$$

por lo que se puede usar el truco de reparametrización para generar muestras de x usando la matriz de covarianza

$$\hat{x} = \mu_\theta(z) + L_\theta(z)^{-1} \varepsilon, \quad \varepsilon \sim N(0, I) \quad (12)$$

donde el inverso de L_θ puede obtenerse fácilmente por sustitución.

3.2. Datos

Para entrenar y validar los modelos se usarán dos bases de datos con imágenes cotidianas e imágenes astronómicas, respectivamente. El primer conjunto corresponde a Fashion-MNIST, el cual consta de 70.000 imágenes de 28×28 píxeles que corresponden a 10 tipos de artículos de moda [22]. Esta base de datos fue diseñada como un reemplazo al tradicional benchmark MNIST, representando un mayor desafío para modelos profundos modernos, pero siendo aún fácilmente interpretable y con un volumen moderado. Por esta razón seleccionamos FMNIST como benchmark para validar conceptualmente la arquitectura previa a su aplicación en imágenes astronómicas. La Figura 1 muestra un ejemplo de cada categoría del dataset FMNIST.

El segundo conjunto de datos proviene del sondeo High-Cadence Transient Survey (HiTS) [8]. HiTS se llevó a cabo entre 2013 y 2015 recuperando alrededor de 100.000 candidatos a transiente astronómico por noche de observación en el telescopio Blanco (4M), en Cerro Tololo, IV Región. En este contexto, un candidato corresponde al brillo residual significativo en la resta entre la imagen

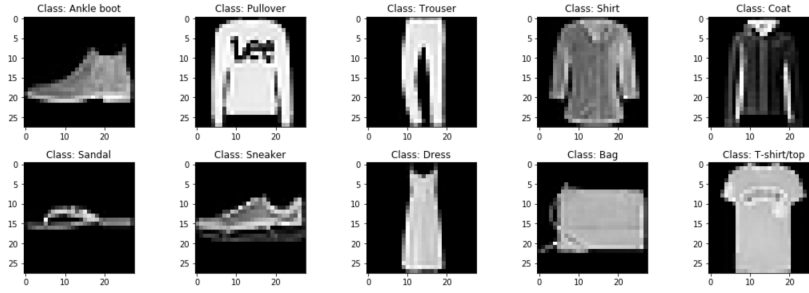


Figura 1. Ejemplo de cada tipo de clase del set de datos Fashion MNIST.

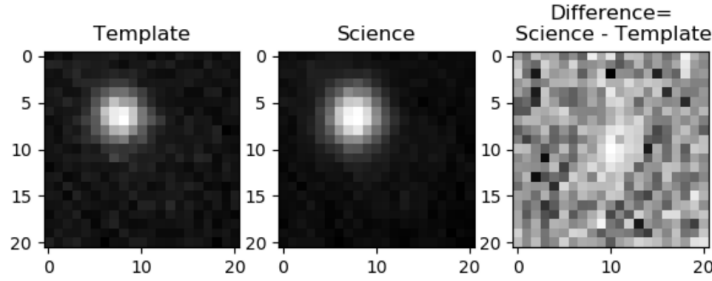


Figura 2. Ejemplo de un candidato a transiente astronómico del set de datos HiTS.

de telescopio, denominada *Science*, y una imagen de referencia capturada con anterioridad, denominada *Template*. En torno al candidato se recorta una zona de 21×21 píxeles como se muestra en la Figura 2.

El conjunto original de candidatos a transiente está dominado por artefactos instrumentales y errores del procedimiento de resta de imágenes. La base de datos fue balanceada artificialmente inyectando ejemplos de transiente real (Supernovae) para entrenar un clasificador Random Forest que fue utilizado durante la campaña de sondeo [8]. La base de datos resultante tiene aproximadamente un millón de candidatos etiquetados como artefacto o transiente real.

3.3. Arquitectura y Experimentos

La base de datos de imágenes se divide en conjuntos de entrenamiento, validación y prueba de forma aleatoria y estratificada en proporción 0.6, 0.25 y 0.15 del total, respectivamente. Con esto se entrena el modelo propuesto con covarianza completa y dos modelos referenciales, uno con covarianza esférica (VAE convencional) y otro con covarianza diagonal, respectivamente. Se usa la misma arquitectura codificadora y la misma cantidad de variables latentes para los tres modelos. Los modelos sólo difieren en el decodificador y en la función de costo utilizada. Todos los modelos se implementan utilizando la librería *PyTorch* [18].

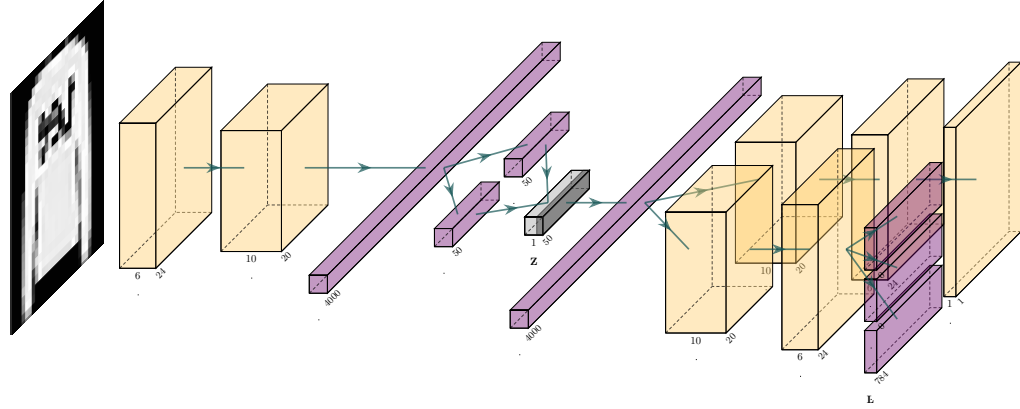


Figura 3. Arquitectura del VAE con covarianza factorizada.

Un esquemático del modelo con covarianza completa se muestra en la Figura 3. El codificador está compuesto por dos capas convolucionales seguidas de dos capas completamente conectadas, las cuales retornan los parámetros μ y σ de la distribución que modela el espacio latente z . Análogamente, el decodificador está compuesto por dos capas completamente conectadas seguidas de sub-redes con dos capas de convolución transpuesta cada una. La primera sub-red retorna la media μ de la distribución generativa. La segunda sub-red retorna la diagonal principal de L , la diagonal asociada al píxel vecino derecho y la diagonal asociada al píxel vecino inferior. Para una imagen de $N \times N$, el píxel derecho del i -ésimo elemento se ubica en $i + 1$, mientras que el píxel inferior se ubica en $i + N$. Todos los demás elementos de L son cero. Otras consideraciones importantes incluyen parametrizar la diagonal principal de L como $\log L$ para forzar la no-negatividad, y forzar que la correlación entre el último elemento de una fila con el primer elemento de la fila siguiente sea cero.

Para entrenar los modelos se utiliza gradiente descendente estocástico con minibatches de tamaño 256 y tasa de aprendizaje adaptativa [13]. La tasa de aprendizaje inicial es 10^{-3} . El entrenamiento se detiene si se detecta que la función de costo no disminuye luego de 20 épocas para evitar el sobreajuste. Los parámetros que alcanzan el menor costo en el conjunto de validación se guardan para realizar la comparación final entre los tres modelos.

Los modelos se comparan cuantitativamente en función de la calidad de generación alcanzada en base al negativo del logaritmo de la verosimilitud $-\mathbb{E}_z[\log p_\theta(x|z)]$ en el conjunto de test en función de la cantidad de variables latentes. Esta métrica es equivalente al error de reconstrucción del modelo.

4. Resultados

En una primera etapa se considera sólo el conjunto de imágenes FMNIST por ser más simple e interpretable. Una vez concluidos los experimentos con FMNIST se procederá a trabajar con la base de datos HiTS. Se entrenaron tres modelos VAE utilizando distintas covarianzas para el decodificador: Esférica o identidad, Covarianza diagonal y Covarianza completa factorizada que corresponde al modelo propuesto. El cuadro 1 muestra el error de reconstrucción en el conjunto de prueba y la época de convergencia. La época varía debido al criterio de early stopping. Todos los modelos se entrenaron usando un espacio latente de tamaño 50. Se puede notar que el error de reconstrucción es más bajo en el modelo propuesto.

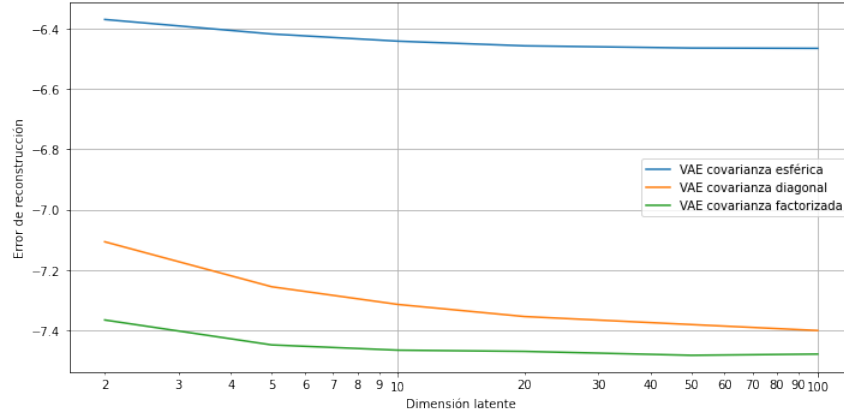
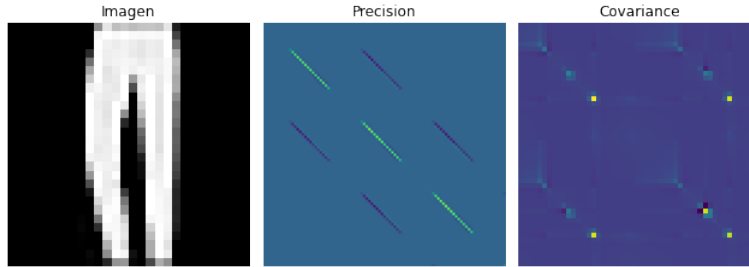
También se estudia la calidad de la reconstrucción en función de la dimensionalidad del espacio latente. La Figura 4 muestra el logaritmo del error de reconstrucción en el conjunto de prueba con un espacio latente que fluctúa entre 5 y 100. Se puede observar que a partir de una dimensión latente de 20, el error de reconstrucción converge para todos los modelos.

La Figura 5 muestra un ejemplo de FMNIST junto a su correspondiente matriz de precisión calculada a partir de su factor de Cholesky L_θ (subfigura central). La matriz de covarianza (subfigura derecha) no es usada durante el entrenamiento del modelo, pero puede calcularse a partir del factor de Cholesky usando la Ec. (11). Por construcción la matriz de precisión es sparse. Sin embargo la matriz de covarianza resultante es considerablemente más densa. De esta forma logramos modelar múltiples correlaciones de forma económica.

Finalmente se analizan las muestras generadas a partir de la distribución decodificadora. Las figuras 6 y 7 muestran dos ejemplos de FMNIST, la media del decodificador y una muestra generada aleatoriamente para cada uno de los tres modelos. Las muestras generadas por el modelo con covarianza esférica, es decir el VAE convencional, son equivalentes a sumar ruido blanco gaussiano sobre la media μ , lo cual es poco realista. La covarianza diagonal permite que los píxeles varíen cada uno en su propia escala logrando una reconstrucción más realista que el caso esférico, pero aun insuficiente. En particular este modelo es capaz de distinguir mejor el objeto del fondo, pero los detalles del objeto siguen cubiertos por ruido blanco. La reconstrucción con covarianza completa generaliza el caso anterior agregando correlaciones entre píxeles. Esto se traduce en detalles mucho más claros y una mejor definición del objeto en cuestión. Sin embargo notamos que en algunos casos se producen artefactos en la forma de píxeles con valores extremos y/o alteraciones en la escala de color con respecto a la imagen original como en el caso de la Figura 6, subfigura de la derecha parte inferior. Esto se debe a errores numéricos durante la inversión del factor de Cholesky. En dichos casos la covarianza asociada no resulta semidefinida positiva, es decir que las restricciones impuestas sobre los valores de L_θ son insuficientes para asegurar una covarianza bien formada.

Cuadro 1. Comparación del error de reconstrucción alcanzado con los tres modelos

Covarianza	$-\mathbb{E}_z[\log p_\theta(x z)]$	Épocas
Esférica	-642	217
Diagonal	-1632	112
Completa factorizada	-1779	116

**Figura 4.** Gráfica comparativa del error de reconstrucción obtenido al entrenar tres modelos con distinta covarianza en la distribución del decodificador en función el tamaño del espacio latente.**Figura 5.** Ejemplo de matriz de precisión y covarianza para una imagen del conjunto de datos Fashion MNIST.

5. Conclusiones

En este trabajo se propuso, implementó y evaluó un autoencoder variacional con covarianza completa en el decodificador levantando el supuesto clásico de independencia entre los atributos de la variable observada. Las pruebas realizadas muestran que el modelo propuesto obtiene un menor error de reconstrucción en la base datos de imágenes FMNIST con respecto a los modelos clásicos, incluso utilizando una menor cantidad de variables latentes. También podemos notar

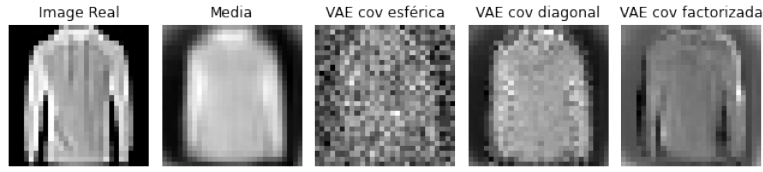


Figura 6. Ejemplo de reconstrucción con 3 modelos diferentes para una imagen del conjunto de datos Fashion MNIST.

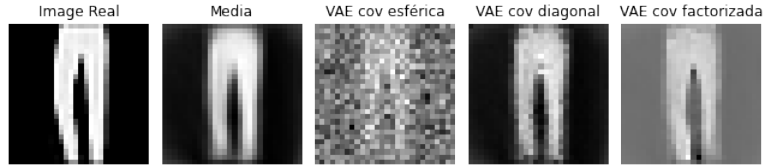


Figura 7. Ejemplo de reconstrucción con 3 modelos diferentes para una imagen del conjunto de datos Fashion MNIST.

que las imágenes generadas por el modelo propuesto son cualitativamente superiores, siendo más realistas y rescatando mejor los detalles finos presentes en los datos.

De estos resultados podemos concluir que el modelo propuesto permite modelar de mejor manera datos con alta correlación como las imágenes. El próximo paso es evaluar el modelo en el conjunto de imágenes astronómicas de HiTS con el fin de comprobar o refutar la hipótesis principal de este trabajo. También resta estudiar en mayor profundidad las condiciones numéricas que han de cumplirse para obtener covarianzas válidas y reconstrucciones libres de artefactos. Una opción basada en primeros principios sería incorporar un prior para el factor de Cholesky, que sea lo suficientemente flexible y que a la vez garantice covarianzas bien formadas. Priors típicamente usados para matrices de covarianza/precisión son la distribución de Wishart inversa y LKJ. También se explorará la implementación de matrices sparse de pytorch que podría dar no sólo mayor eficiencia sino también mejor precisión numérica.

Referencias

1. Baldi, P.: Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning. pp. 37–49 (2012)
2. Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: A review for statisticians. *Journal of the American statistical Association* **112**(518), 859–877 (2017)
3. Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P.A., Maureira, J.C.: Deep-hits: Rotation invariant convolutional neural network for transient detection. *The Astrophysical Journal* **836**(1), 97 (2017)

4. Dahmen, J., Keysers, D., Pitz, M., Ney, H.: Structured covariance matrices for statistical image object recognition. In: *Mustererkennung 2000*, pp. 99–106. Springer (2000)
5. Dempster, A.P.: Covariance selection. *Biometrics* pp. 157–175 (1972)
6. Dorta, G., Vicente, S., Agapito, L., Campbell, N.D., Simpson, I.: Structured uncertainty prediction networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5477–5485 (2018)
7. Dorta, G., Vicente, S., Agapito, L., Campbell, N.D., Simpson, I.: Training vaes under structured residuals. *arXiv preprint arXiv:1804.01050* (2018)
8. Förster, F., Maureira, J.C., San Martín, J., Hamuy, M., Martínez, J., Huijse, P., Cabrera, G., Galbany, L., De Jaeger, T., González-Gaitán, S., et al.: The high cadence transient survey (hits). i. survey design and supernova shock breakout constraints. *The Astrophysical Journal* **832**(2), 155 (2016)
9. Goodfellow, I., Bengio, Y., Courville, A.: *Deep learning*. MIT press (2016)
10. Hoffman, M.D., Blei, D.M., Wang, C., Paisley, J.: Stochastic variational inference. *The Journal of Machine Learning Research* **14**(1), 1303–1347 (2013)
11. Ivezić, Ž., Kahn, S.M., Tyson, J.A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S.F., Andrew, J., et al.: Lsst: from science drivers to reference design and anticipated data products. *The Astrophysical Journal* **873**(2), 111 (2019)
12. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European conference on computer vision*. pp. 694–711. Springer (2016)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013)
15. Krizhevsky, A., Hinton, G.E.: Using very deep autoencoders for content-based image retrieval. In: *ESANN*. vol. 1, p. 2. Citeseer (2011)
16. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3431–3440 (2015)
17. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional auto-encoders for hierarchical feature extraction. In: *International conference on artificial neural networks*. pp. 52–59. Springer (2011)
18. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: *Advances in neural information processing systems*. pp. 8026–8037 (2019)
19. Perlmutter, S., et al.: Supernovae, dark energy, and the accelerating universe. *Physics today* **56**(4), 53–62 (2003)
20. Pu, Y., Gan, Z., Hénao, R., Yuan, X., Li, C., Stevens, A., Carin, L.: Variational autoencoder for deep learning of images, labels and captions. In: *Advances in neural information processing systems*. pp. 2352–2360 (2016)
21. Rezende, D.J., Mohamed, S., Wierstra, D.: Stochastic backpropagation and approximate inference in deep generative models. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume 32*. pp. II–1278 (2014)
22. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747* (2017)