

Autoencoders

Denoising Autoencoders - Sparse Autoencoders

AQUISE ANGELA, CASTILLO KEMELY y JARA ARTURO

Universidad Nacional de San Agustín

1 Autoencoders

- Definición
- Arquitectura
- Aplicación de Autoencoders

2 Denoising Autoencoders

- Definición
- DAE
- MODIFIED DENOISING OBJECTIVE
- Conclusión

3 Sparse Autoencoders

- Introducción
- Kullback-Liebler

Autoencoder Definición

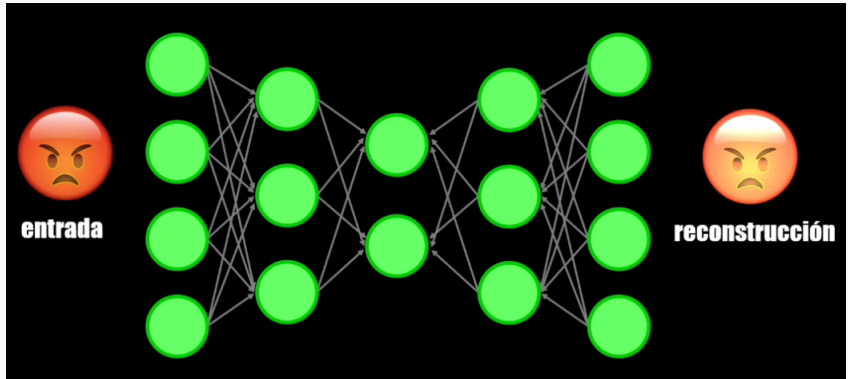


Figura 1: Esquema de un Autoencoder

Función Coder: $h=f(x)$

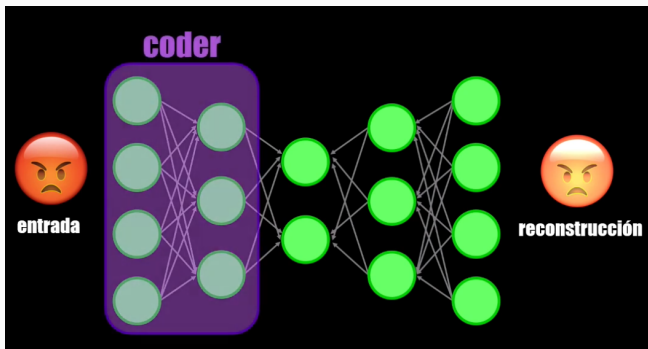


Figura 2: Endocer

Bottleneck

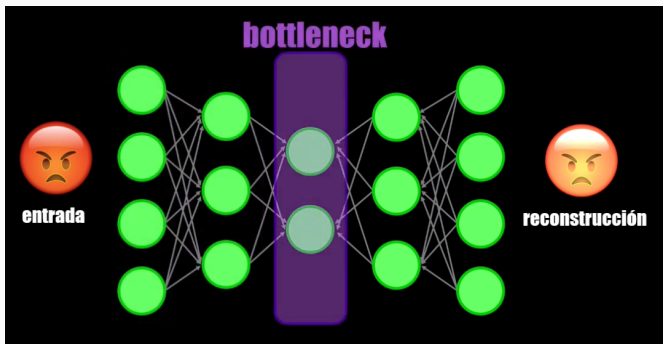


Figura 3: Bottleneck

Función Decoder: $r=g(h)$

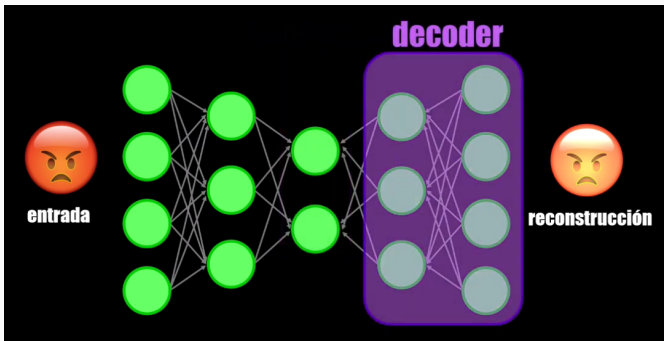


Figura 4: Decoder

Función: $g(f(x))=r$

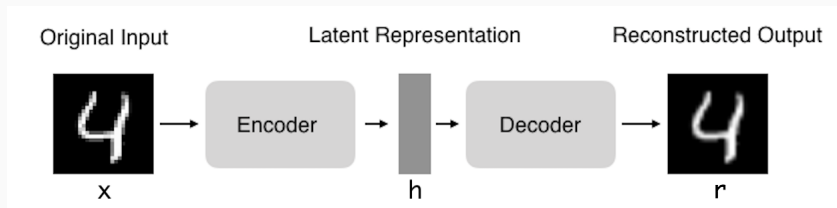


Figura 5: Esquema de un Autoencoder



Figura 9: Empresa que utiliza Autoencoders

1. Detección de anomalías
 - Empresas como Paypal.
2. Comprensión de información
 - Características más importantes.
3. Generación de imágenes
 - Tareas de diseño, nuevos datasets, etc.

Recordatorio: Los autoencoders son redes neuronales que se usan comúnmente para la selección y extracción de características.

Problemas: Cuando hay más nodos en la *hidden layer* que en los *inputs*, la red se arriesga a aprender la llamada **Identity Function** (*función de identidad*), también conocida como **Null Function** (*función nula*)

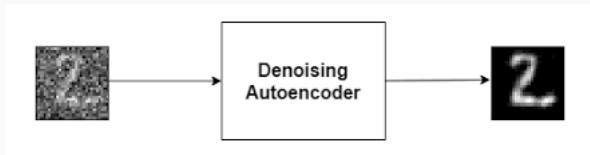


Figura 10

Denoising Autoencoders DAE

- El *denoising autoencoder* se entrena minimizando el error de los datos y una reconstrucción obtenida a partir de una versión corrupta con ruido.
- Como consecuencia el autoencoder aprende a "limpiar" los datos

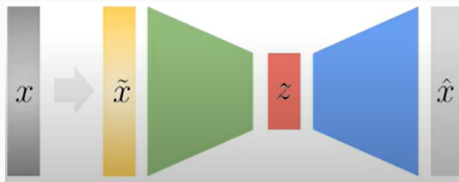


Figura 11

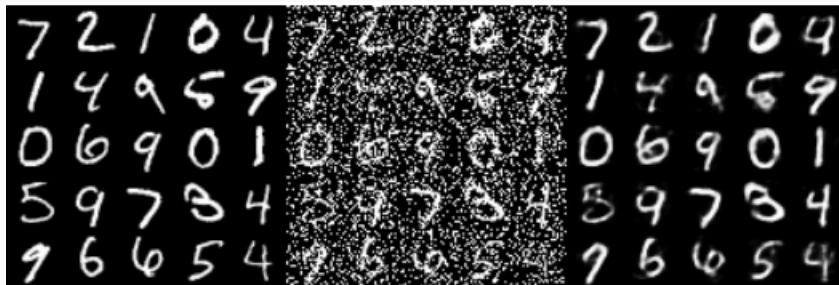


Figura 12. DAE

Denoising Autoencoder DAE

- Una corrupción típicamente usada es agregar ruido blanco a los datos.
- El autoencoder se volverá robusto a este tipo de ruido.
- En este caso la función de costo que se está optimizando es:

$$\min \mathcal{L}(\theta, \phi) = \mathbb{E}_{\tilde{x}_i \sim \mathcal{N}(x_i, \sigma^2)} [L(x_i, f_{\theta}(g_{\phi}(\tilde{x}_i)))]$$

- La varianza del ruido es un hiper parámetro del modelo
- Distintas corrupciones le otorgarán distintas propiedades al autoencoder

El objetivo del DAE es aprender una composición encoder-decoder determinista

$$f(.) = G^{-1}(G(.))$$

tal que para

$$Y = f(\tilde{X})$$

la cantidad $E[D(Y, X)]$ es minimizado.

El proceso de corrupción conduce al aprendizaje de transformaciones no triviales de la entrada en particular para los casos demasiado completos en los que se puede aprender el mapeo de identidad si no hay la capacidad adecuada mecanismo de control en su lugar [3].

$$\underset{\theta \in \Theta}{\text{minimize}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\hat{X}_i} [D(f_{\theta}(\hat{X}_i), \mathbf{x}_i)],$$

Figure: Parámetro aprendizaje del DAE se puede formular como el siguiente problema de optimización

- En este enfoque, solo la dimensionalidad y el rango del encoder pueden controlarse modificando su arquitectura.
- Por lo tanto, existe la posibilidad de que el encoder podría estar aprendiendo un mapa de características, para el cual algunas dimensiones aún tienen los efectos de perturbar la entrada original.

Para hacer cumplir explícitamente la robustez en la fase de codificación, podemos medir la cantidad de distorsión en el encoder comparando los valores de característica resultantes para la entrada original no corrupta con los valores de característica obtenidos al codificar su contraparte corrupta.

$$\underset{\theta \in \Theta}{\text{minimize}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\hat{X}_i} [D(f_{\theta}(\hat{X}_i), \mathbf{x}_i)] + \frac{1}{\lambda} \mathbb{E}_{\hat{X}_i} [D_{\text{enc}}(G(\hat{X}_i), G(\mathbf{x}_i))],$$

Figure:

(λ) es un parámetro de compensación

Denoising Autoencoders MODIFIED DENOISING OBJECTIVE

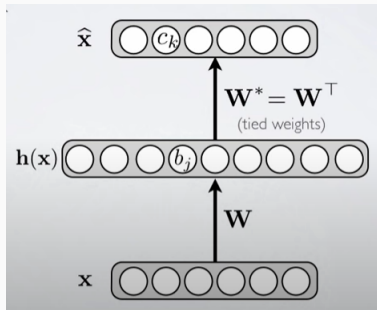


Figura 13. AE

- La capa oculta está demasiado completa si es mayor que la capa de entrada.

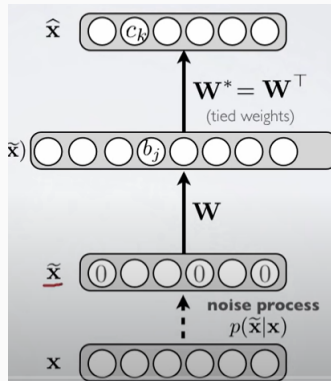
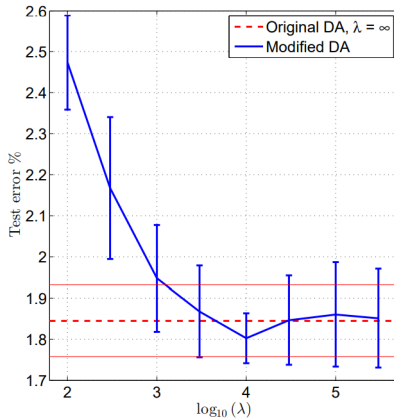


Figura 14. DAE



Para diferentes valores de λ

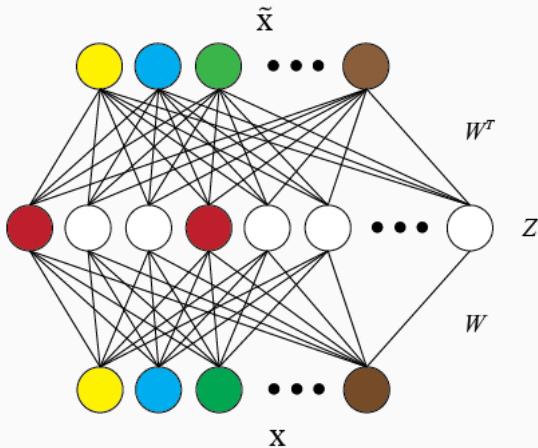


Figura 15: Sparse autoencoder [5]

- Si existe una restricción adicional en la red no es necesario un cuello de botella.
- La restricción particular que estamos considerando es **sparsity**.
- Al restringir las activaciones de los nodos ocultos, de hecho estamos reduciendo el número de neuronas que se activan.

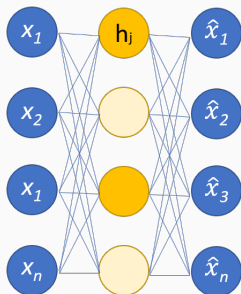
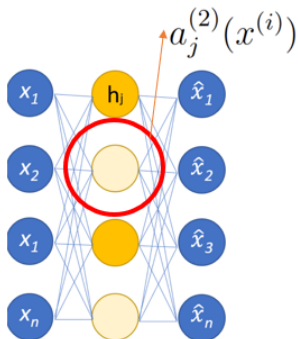


Figura 16: Un simple sparse auto encoder de una sola capa con el mismo número de entradas (x), salidas y nodos ocultos (h). Los nodos amarillos más oscuros se activan mientras que los nodos más claros están restringidos.



Activación promedio de la unidad j

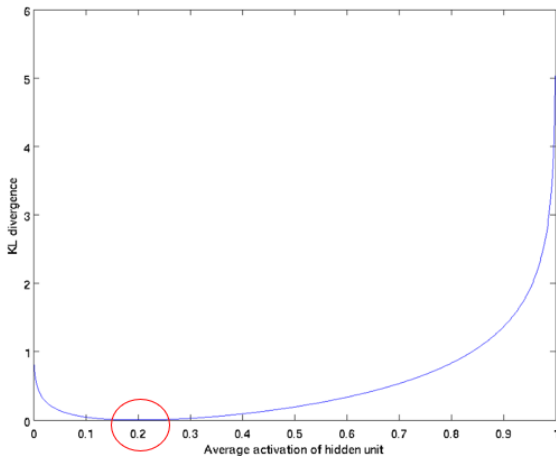
$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m \left[a_j^{(2)}(x^{(i)}) \right]$$

Restringir las neuronas para que permanezcan inactivas la mayor parte del tiempo, es decir, con una salida cercana a 0.

- Para crear una restricción, tratamos de imponer que la activación promedio \bar{p} sea igual a un parámetro específico cercano a cero (a menudo 0.05). Para satisfacer esa restricción, la activación oculta debe estar cerca de 0 según sea necesario.

$$\sum_{j=1}^{s_2} \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j}.$$

$$\sum_{j=1}^{s_2} \text{KL}(\rho || \hat{\rho}_j),$$



$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho || \hat{\rho}_j),$$

- En general, al agregar esta penalización a la función de costo, es posible regularizar la salida de un codificador automático estándar o un codificador automático que contiene relativamente muchas unidades ocultas.

- [1] R. Cañadas.
Qué son los autoencoders, 8 2021.
- [2] R. G. Dor Bank, Noam Koenigstein.
Autoencoders.
2021.
- [3] L. G. Sanchez Giraldo.
Revisiting denoising auto-encoders.
2017.
- [4] M. Sotaquirá.
Autoencoders: explicación y tutorial en python, 5 2019.
- [5] E. Wilkinson.
Deep learning: Sparse autoencoders., 2018.