

Autoencoders

PhD(e). Jonnatan Arias Garcia – jonnatan.arias@utp.edu.co –
jariasg@uniquindio.edu.co

PhD. David Cardenas peña - dcardenas@utp.edu.co

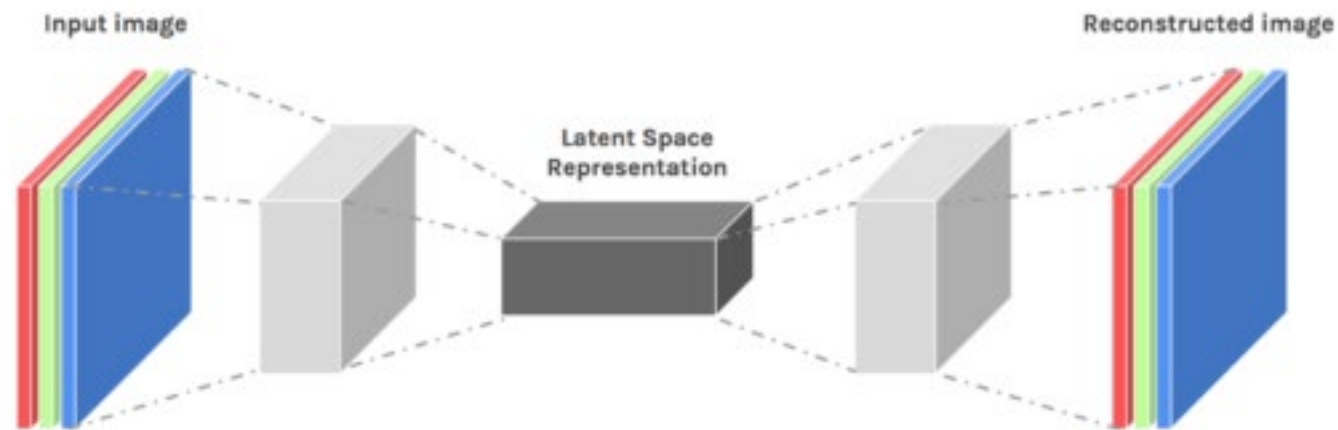
PhD. Hernán Felipe Garcia - hernanf.garcia@udea.edu.co

Autoencoders

- El aprendizaje supervisado utiliza etiquetas explícitas/salida correcta para entrenar una red.
 - Por ejemplo, clasificación de imágenes.
- El aprendizaje no supervisado se basa únicamente en datos.
 - Por ejemplo, embeddings de palabras CBOW y skip-gram: la salida se determina implícitamente a partir del orden de las palabras en los datos de entrada.
 - El punto clave es producir un embedding útil de palabras.
 - El embedding codifica la estructura, como la similitud de palabras y algunas relaciones.
 - Todavía es necesario definir una pérdida: esta es una supervisión implícita.

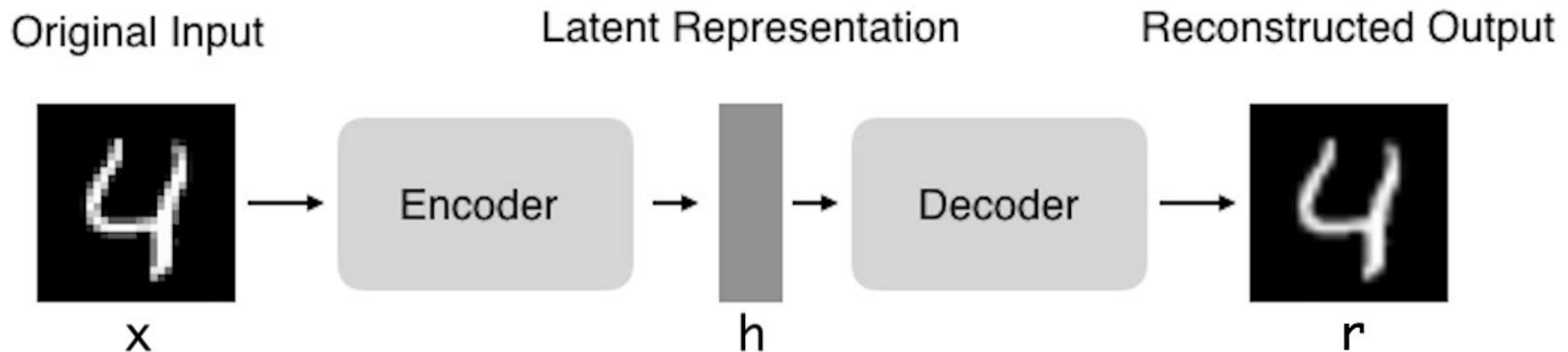
Autoencoders

- Los autocodificadores están diseñados para reproducir su entrada, especialmente para imágenes.
 - El punto clave es reproducir la entrada de una codificación aprendida.



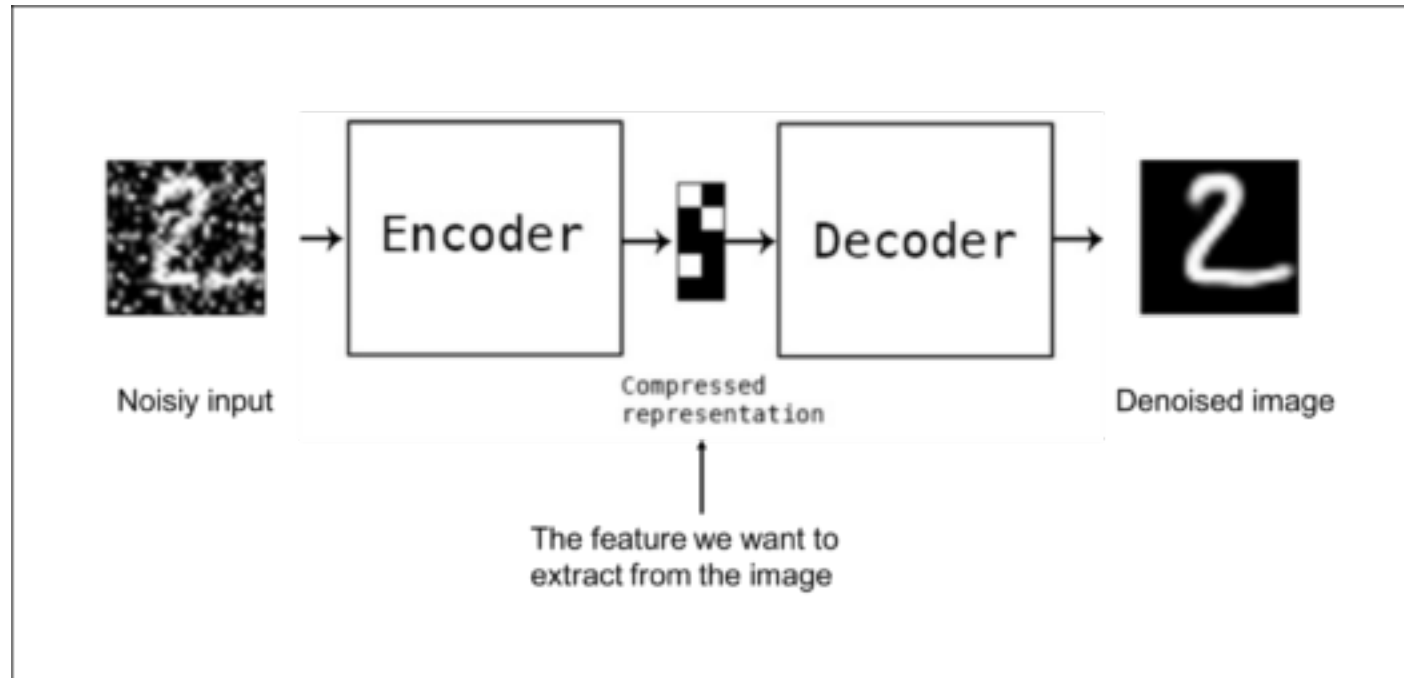
Autoencoders: Estructura

- Codificador (Encoder): comprime la entrada en un espacio latente de dimensiones generalmente más pequeñas. $h = f(x)$
- Decodificador (Decoder): reconstruye la entrada a partir del espacio latente. $r = g(f(x))$ con r lo más cerca posible de x



Autoencoders: Aplicaciones

- Eliminación de ruido: ingrese imagen limpia + ruido y entrene para reproducir la imagen limpia.



Autoencoders: Aplicaciones

- Coloración de imágenes: ingrese blanco y negro y entrene para producir imágenes en color



Autoenconders: Aplicaciones

- Remover marcas de agua



Autoencoders: Propiedades

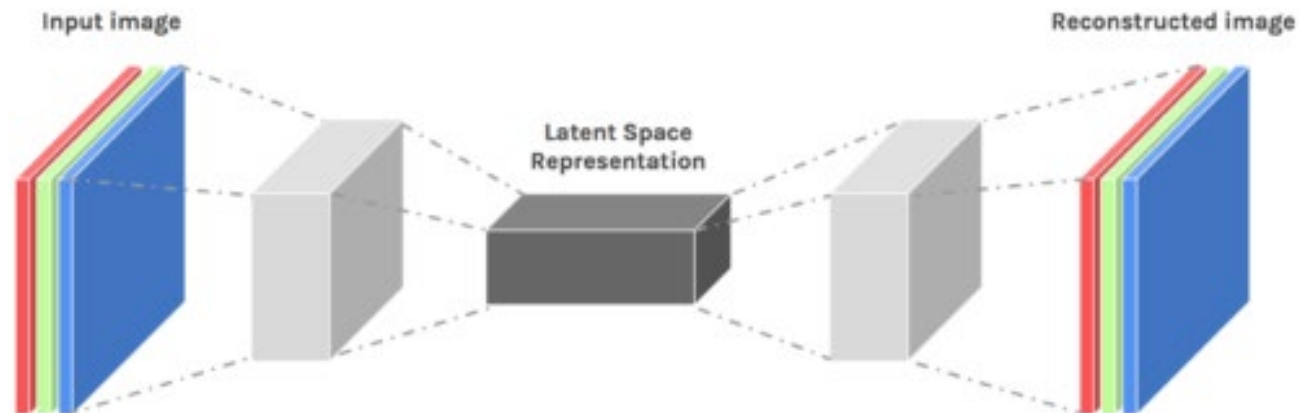
- **Datos específicos:** los codificadores automáticos solo pueden comprimir datos similares a los que se han entrenado.
- **Con pérdidas (Lossy):** Las salidas descomprimidas se degradarán en comparación con las entradas originales.
- **Aprendizaje automáticamente a partir de ejemplos:** Es fácil entrenar instancias especializadas del algoritmo que funcionarán bien en un tipo específico de entrada.

Autoencoders: Capacidad

- Al igual que con otras NN, el sobreajuste es un problema cuando la capacidad es demasiado grande para los datos.
- Los autocodificadores abordan esto a través de una combinación de:
 - Capa de cuello de botella: menos grados de libertad que en las salidas posibles.
 - Entrenamiento para eliminar el ruido.
 - Dispersión a través de la regularización.
 - Penalización contractiva.

Autoencoders: Bottleneck layer

- Supongamos que las imágenes de entrada son **$n \times n$** y el espacio latente es **$m < n \times n$** .
- Entonces el espacio latente no es suficiente para reproducir todas las imágenes.
- Necesita aprender una codificación que capture las características importantes de los datos de entrenamiento, suficientes para una reconstrucción aproximada.



Bottleneck layer en Keras

- `input_img = Input(shape=(784,))`
- `encoding_dim = 32`
- `encoded = Dense(encoding_dim, activation='relu')(input_img)`
- `decoded = Dense(784, activation='sigmoid')(encoded)`
- `autoencoder = Model(input_img, decoded)`
- Maps 28x28 images into a 32 dimensional vector.
- Can also use more layers and/or convolutions.



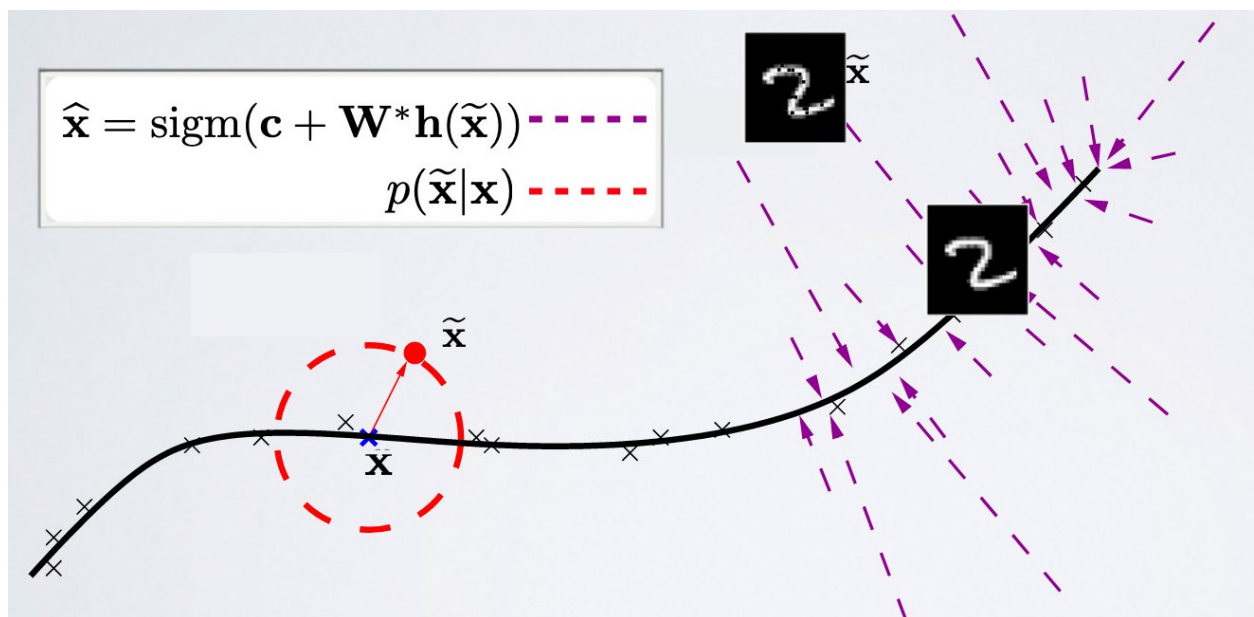
Autoencoders: Denoising

- El autoencoder básico se entrena para minimizar la pérdida entre x y la reconstrucción $g(f(x))$.
 - Los autoencoders de eliminación de ruido se entrenan para minimizar la pérdida entre x y $g(f(x+w))$, donde w es ruido aleatorio.
 - Las mismas arquitecturas posibles, diferentes datos de entrenamiento.
- Kaggle tiene un conjunto de datos sobre documentos dañados.



Autoencoders: Denoising

- Los autoencoders de eliminación de ruido no pueden simplemente memorizar la relación de entrada y salida.
- De forma intuitiva, un autoencoder de eliminación de ruido aprende una proyección de una vecindad de nuestros datos de entrenamiento en los datos de entrenamiento.



Sparse Autoencoders

- Construye una función de pérdida para penalizar las activaciones dentro de una capa.
- Normalmente regularizar los pesos de una red, no las activaciones.
- Los nodos individuales de un modelo entrenado que se activan dependen de los datos.
 - Diferentes entradas darán como resultado activaciones de diferentes nodos a través de la red.
- Active selectivamente las regiones de la red en función de los datos de entrada.

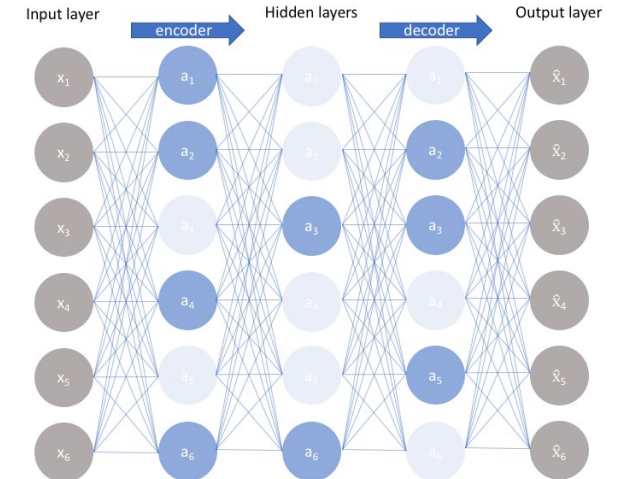
Sparse Autoencoders

- Construya una función de pérdida para penalizar las activaciones de la red.
 - **Regularización L1:** Penaliza el valor absoluto del vector de activaciones a en la capa h para observación l

$$\mathcal{L}(x, \hat{x}) + \lambda \sum_i |a_i^{(h)}|$$

- **Divergencia KL:** Utilice la entropía cruzada entre la activación media y la activación deseada

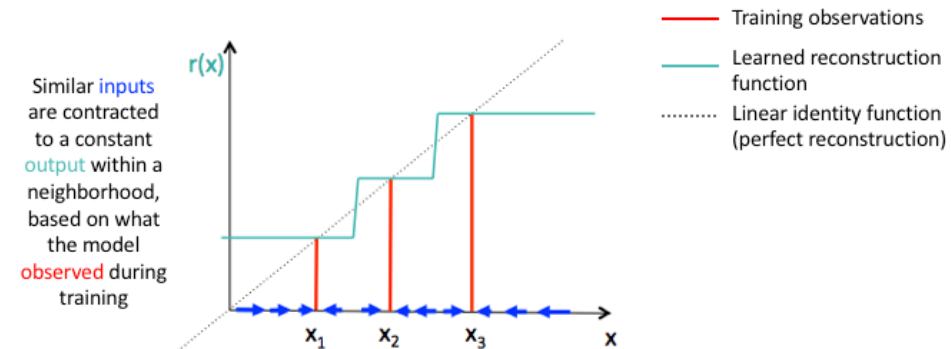
$$\mathcal{L}(x, \hat{x}) + \sum_j KL(\rho || \hat{\rho}_j)$$



Autoencoders Contractivos

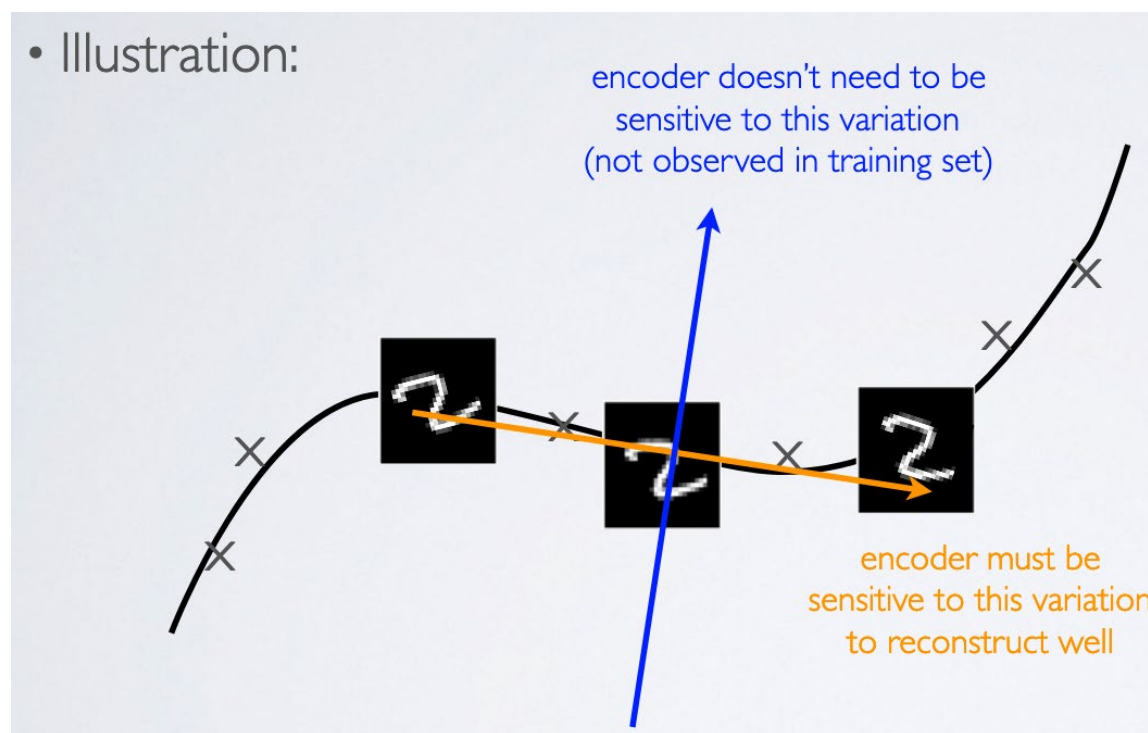
- Organice entradas similares para que tengan activaciones similares.
 - Es decir, la derivada de las activaciones de la capa oculta es pequeña con respecto a la entrada.
- Los autoencoders de eliminación de ruido hacen que la función de reconstrucción (codificador + decodificador) resista pequeñas perturbaciones de la entrada
- Los autoencoders contractivos hacen que la función de extracción de características (es decir, el codificador) resista las perturbaciones infinitesimales de la entrada.

$$\mathcal{L}(x, \hat{x}) + \lambda \sum_i \left\| \nabla_x a_i^{(h)}(x) \right\|^2$$



Autoencoders Contractivos

- Los autocodificadores contractivos hacen que la función de extracción de características (es decir, el codificador) resista las perturbaciones infinitesimales de la entrada.



Autoencoders Contractivos

- Tanto el autoencoders (AE) de eliminación de ruido como el autoencoders contractivo pueden funcionar bien
 - Ventaja de eliminar el ruido del codificador automático: más simple de implementar, requiere agregar una o dos líneas de código al codificador automático normal, sin necesidad de calcular Jacobian de la capa oculta
 - Ventaja del AE contractivo: el gradiente es determinista, puede usar optimizadores de segundo orden (gradiente conjugado, LBFGS, etc.) - podría ser más estable que el AE de eliminación de ruido, que utiliza un gradiente muestreado.
- Para obtener más información sobre los AE's contractivos: }
 - Autocodificadores contractivos: invariancia explícita durante la extracción de características. Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot y Yoshua Bengio, 2011.