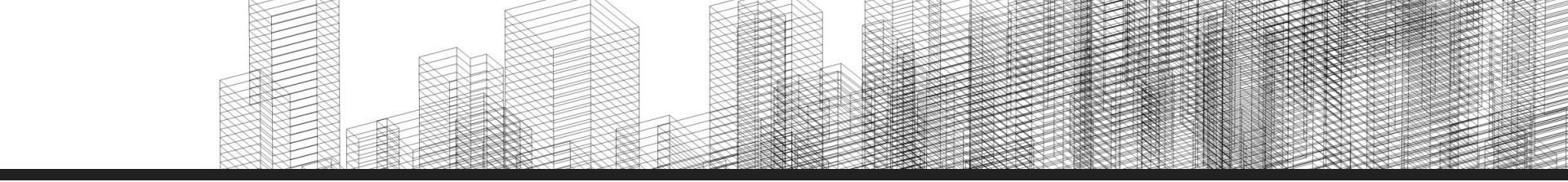
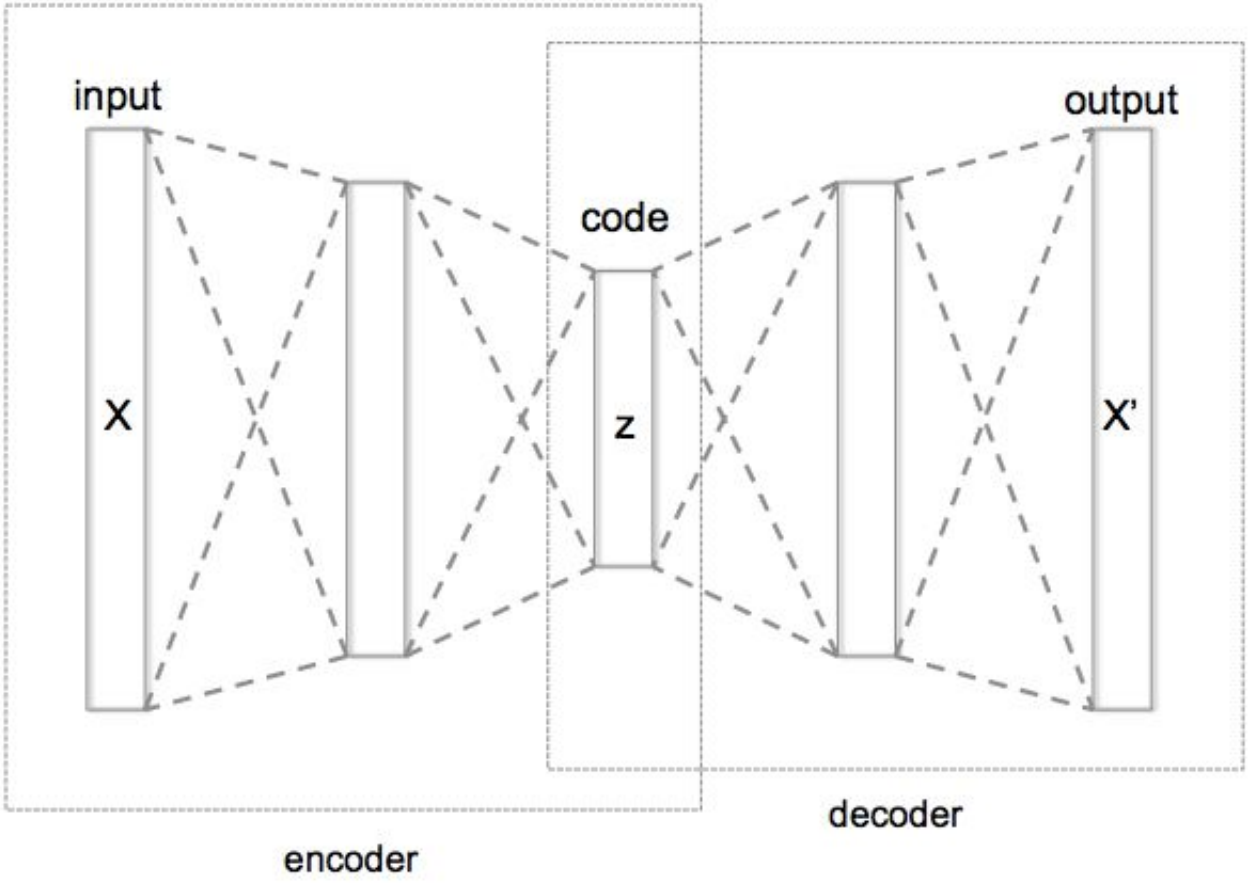


Autoencoders

Deep Learning Book Review
Ch 14 (Part 1)
By Laura Montoya, Founder & CEO @ Accel.AI

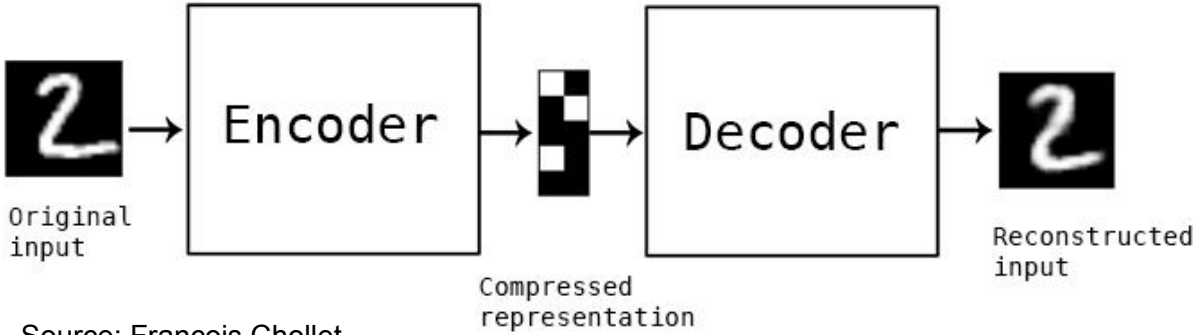
What is an autoencoder?

An autoencoder is a feedforward neural network that is trained to copy its input to its output.

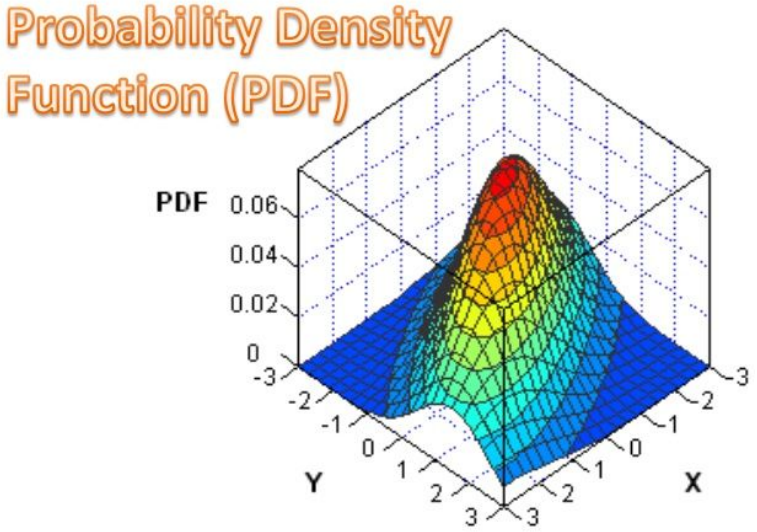


Feature Extraction Engine

Aim of Autoencoder is to learn a representation (probability density function) for a set of data, which is useful to figure out the underlying structure of the data set.

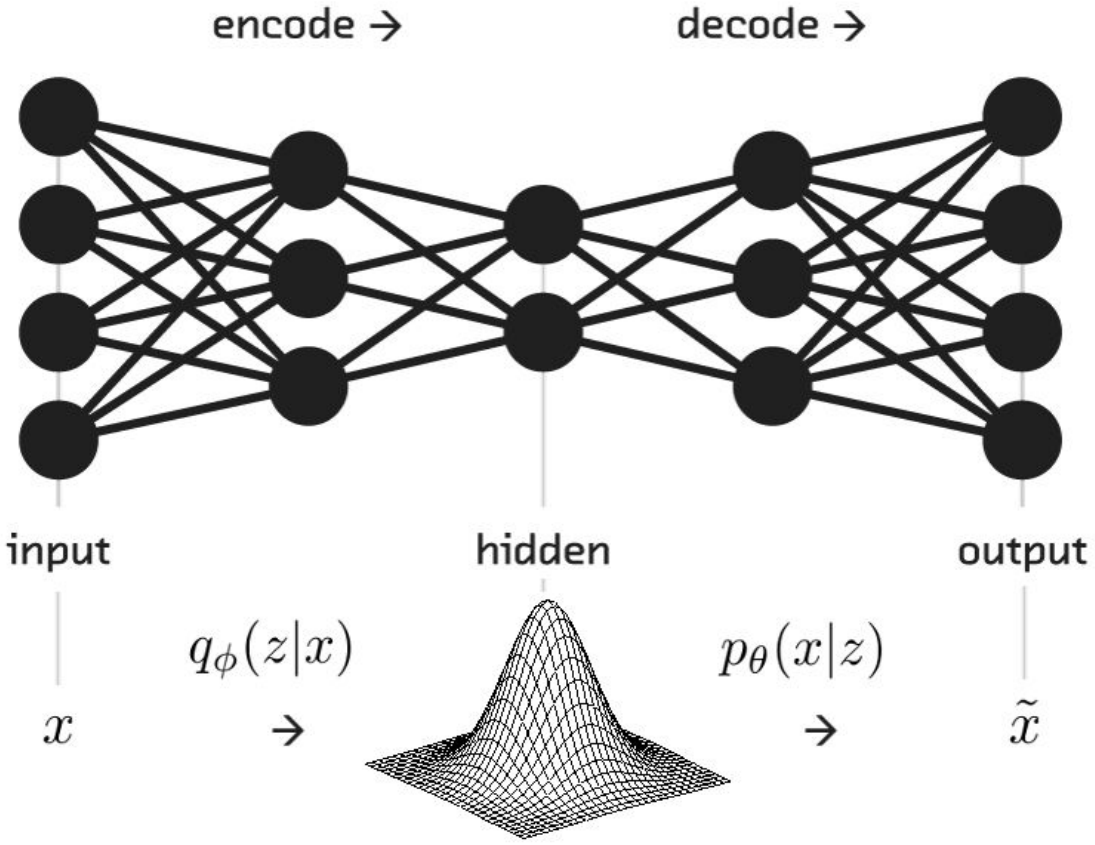


Source: Francois Chollet



Architecture

1. Encoder
2. Hidden (Bottle Neck) Layer
3. Decoder



Encoder

Where h is feature vector or representation or code computed from x.

$$h^{(t)} = f_\theta(x^{(t)}, \{x^{(1)}, \ldots, x^{(T)}\})$$

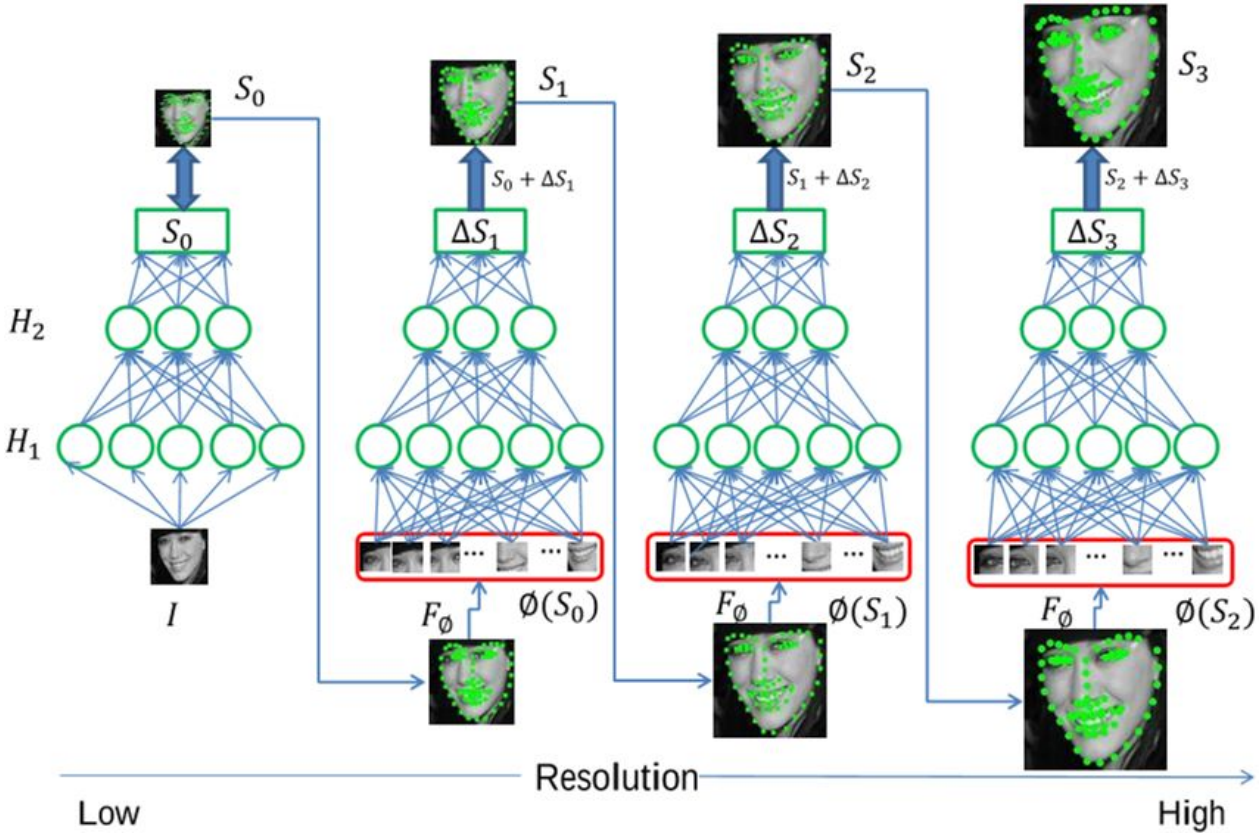
Decoder

Maps from feature space back into input space, producing a reconstruction.

$$r = g_\theta(h)$$

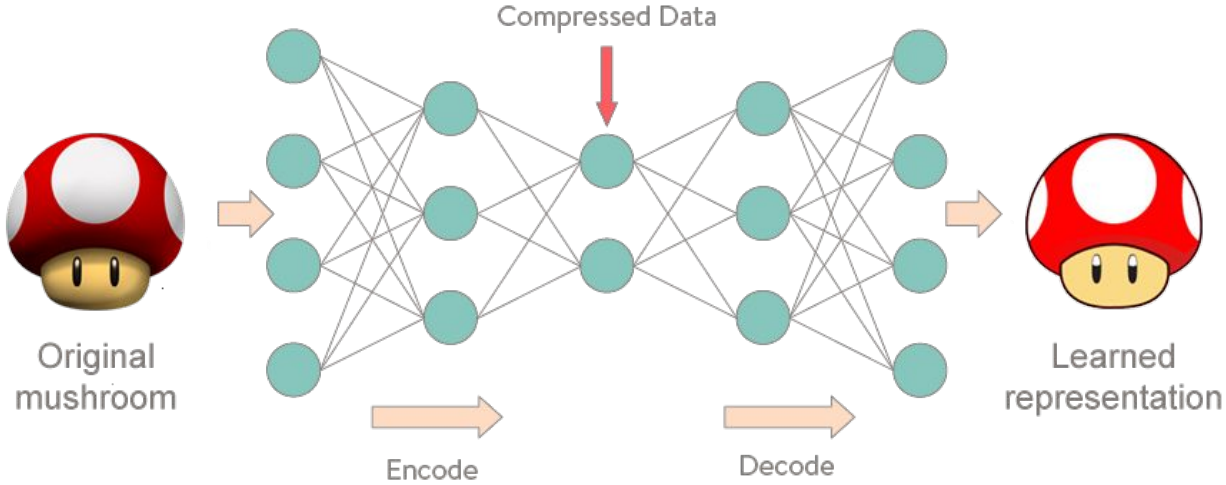
Data Specific

They will only be able to compress data similar to what they have been trained on.



Undercomplete Autoencoders

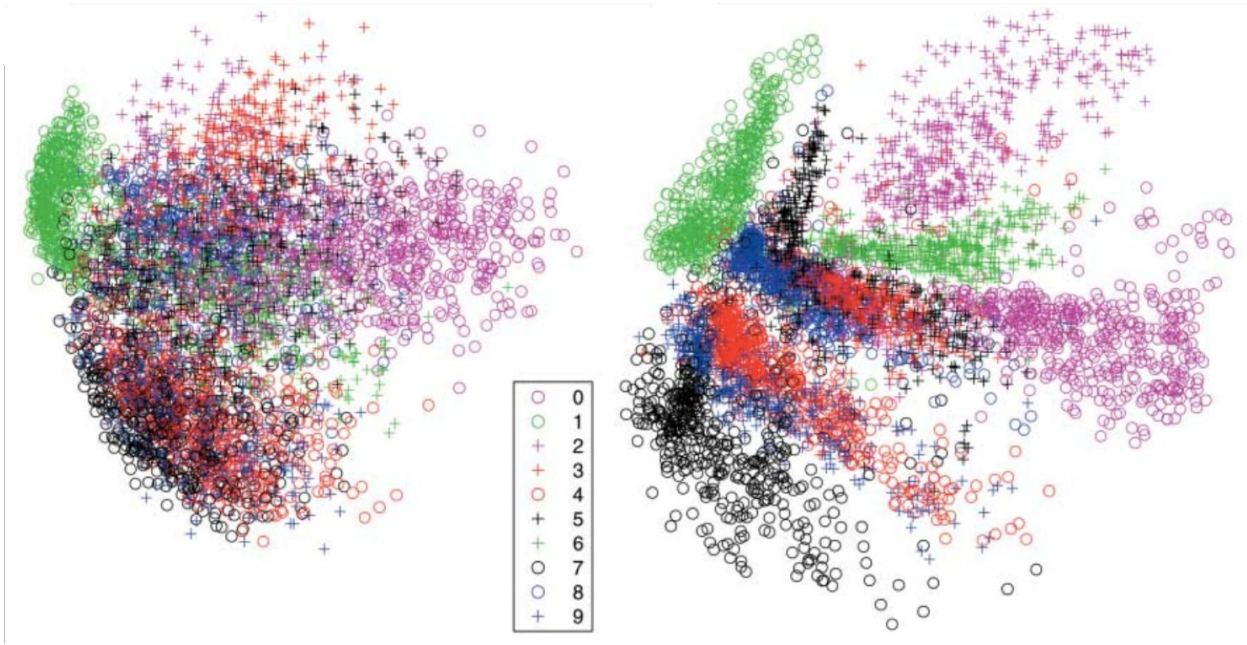
Learning process is described as minimizing a loss function.



$$L(\boldsymbol{x}, g(f(\boldsymbol{x})))$$

AE vs PCA

Undercomplete autoencoder learns to span the same subspace as Principle Component Analysis.



Regularized Autoencoders

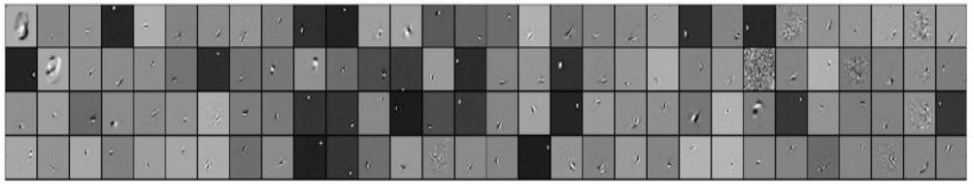
Use a loss function that encourages the model to have other properties besides the ability to copy its input to its output.

- Sparsity of the representation
- Smallness of the derivative of the representation
- Robustness to noise or to missing input

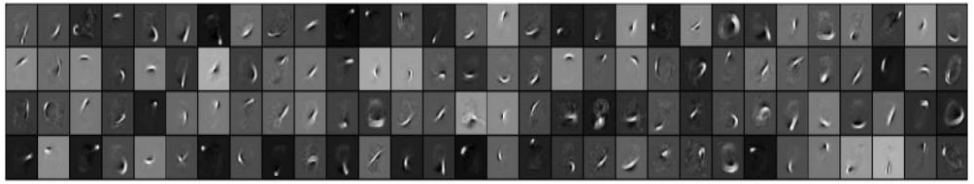
Sparse Autoencoders

Used to learn features for another task such as classification.

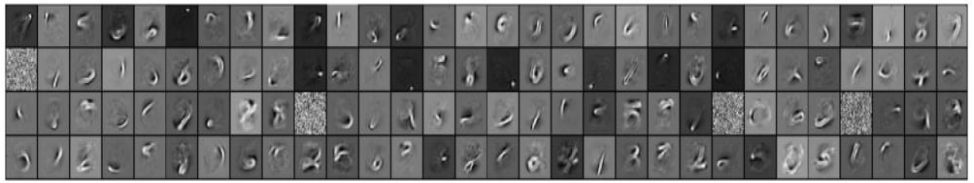
$$L(\boldsymbol{x}, g(f(\boldsymbol{x}))) + \Omega(\boldsymbol{h})$$



(a) $k = 70$

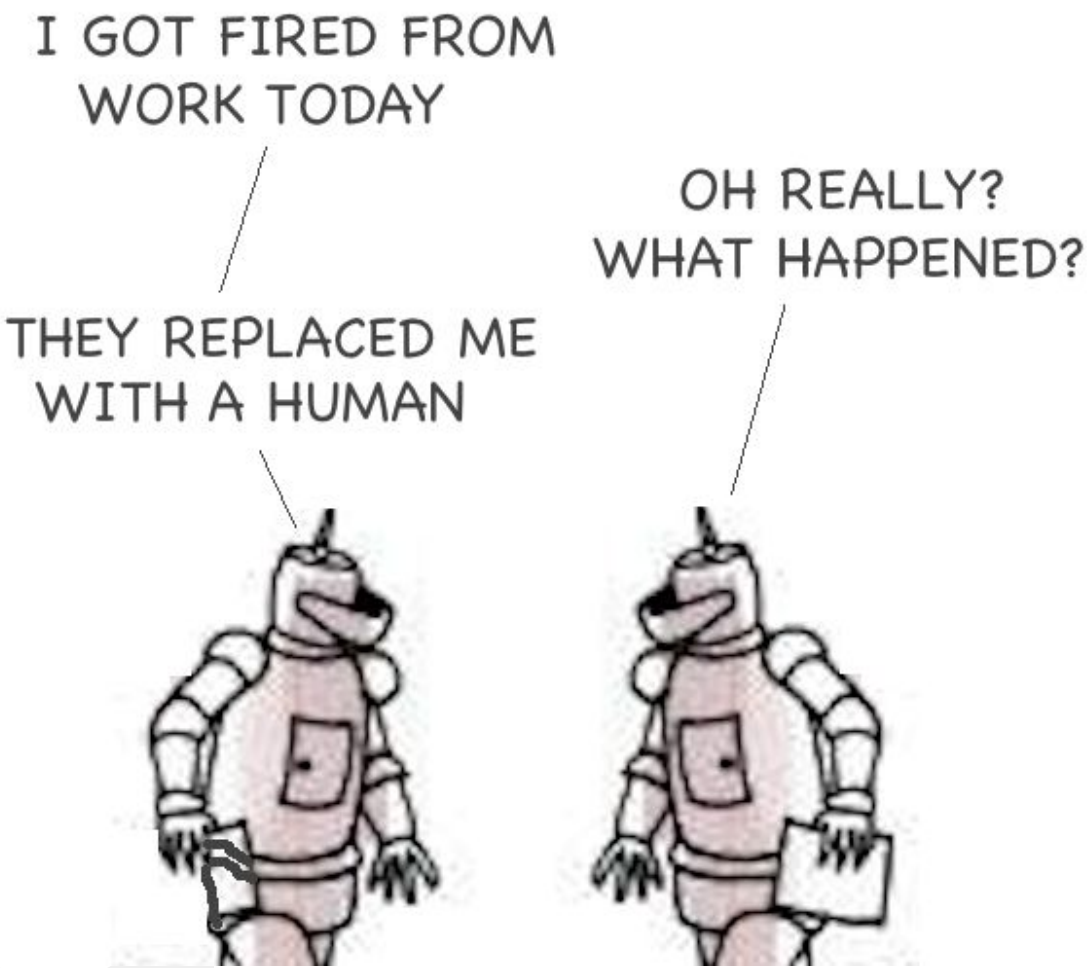


(b) $k = 40$



(c) $k = 25$

(d) $k = 10$


I GOT FIRED FROM WORK TODAY
THEY REPLACED ME WITH A HUMAN
OH REALLY? WHAT HAPPENED?