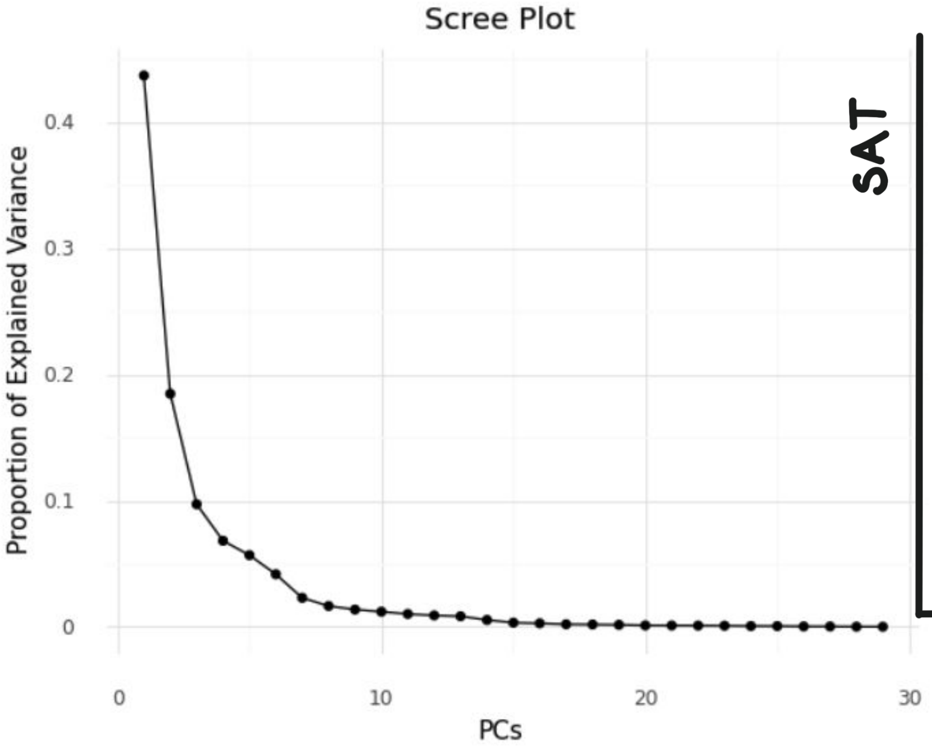
Autoencoders

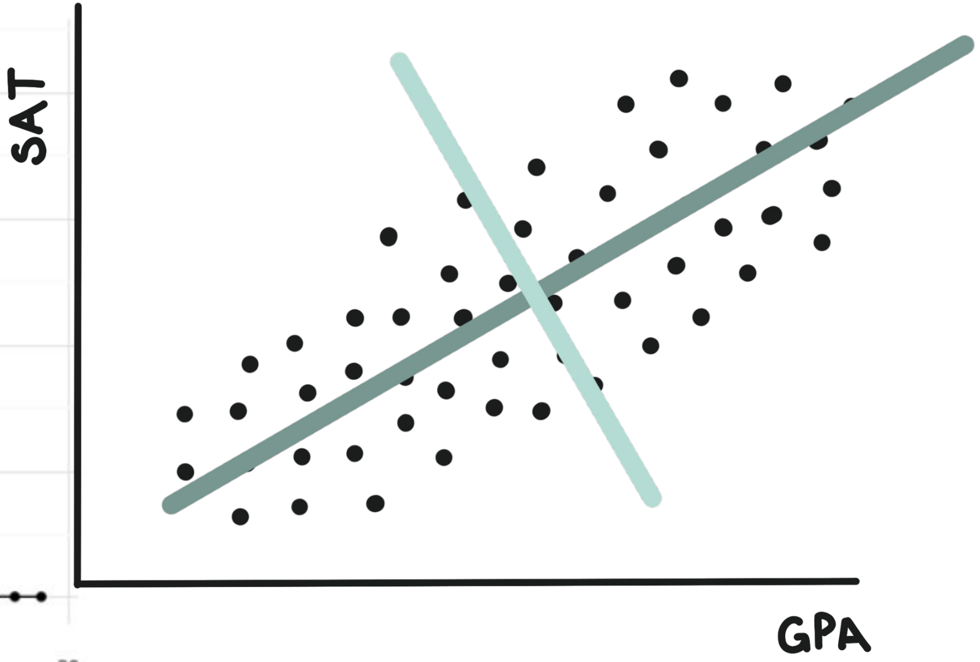
Dr. Chelsea Parlett-Pelleriti

Outline

- PCA
- Architecture
- Nonlinear Encoders/Decoders
- Regularization: Penalization
- Regularization: Penalizing Derivatives
- Regularization: Denoising
- Convolutional Autoencoders
- Example: Image Compression

PCA Review


Scree Plot
Proportion of Explained Variance
0.4
0.3
0.2
0.1
0
0
10
20
30
PCs


SAT
GPA

PCA and Lossy Compression

Original Faces

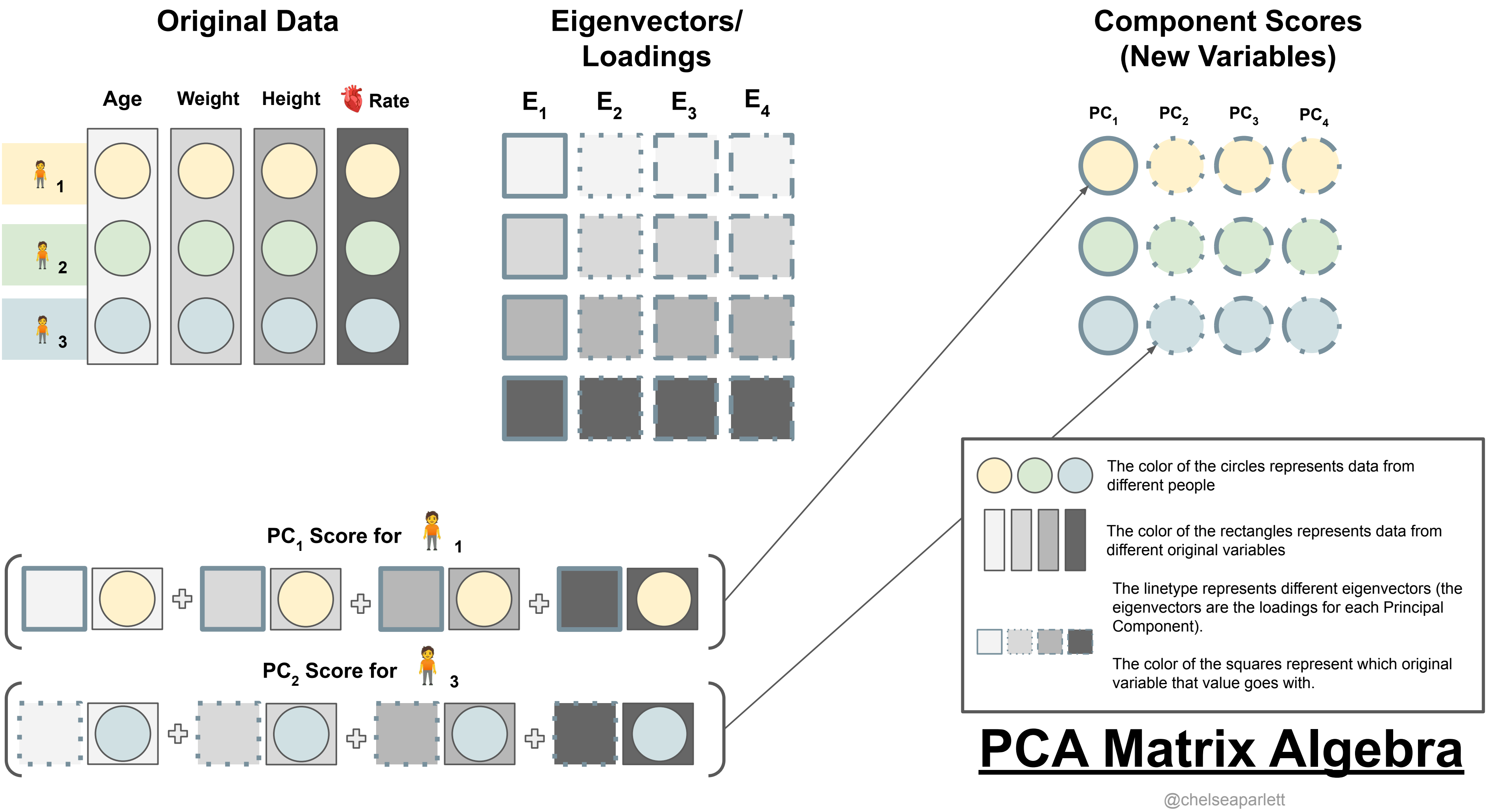
100 Principal Components

10 Principal Components

PCA and Linearity

PC's are **linear combinations** of the original variables

$$w^T x$$

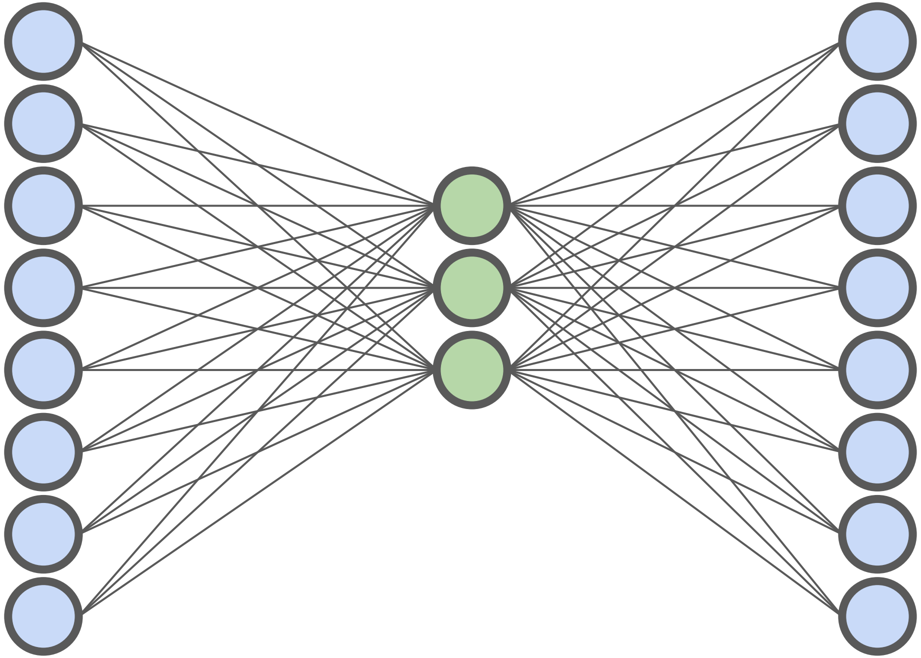


Original Data
Eigenvectors/ Loadings
Component Scores (New Variables)
Age
Weight
Height
Rate
1
2
3
E_1
E_2
E_3
E_4
PC_1
PC_2
PC_3
PC_4
PC_1 Score for 1
PC_2 Score for 3
The color of the circles represents data from different people
The color of the rectangles represents data from different original variables
The linetype represents different eigenvectors (the eigenvectors are the loadings for each Principal Component).
The color of the squares represent which original variable that value goes with.
PCA Matrix Algebra
@chelseaparlett

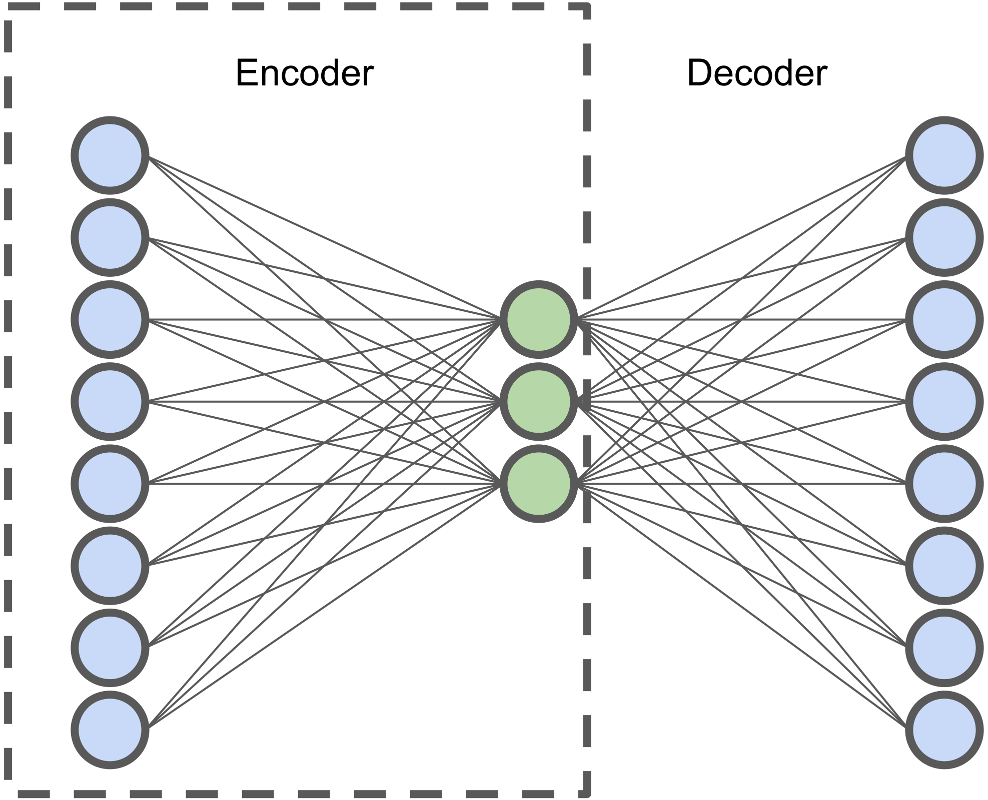
What if we had NON-Linear PCA?


AUTOENCODERS
imgflip.com

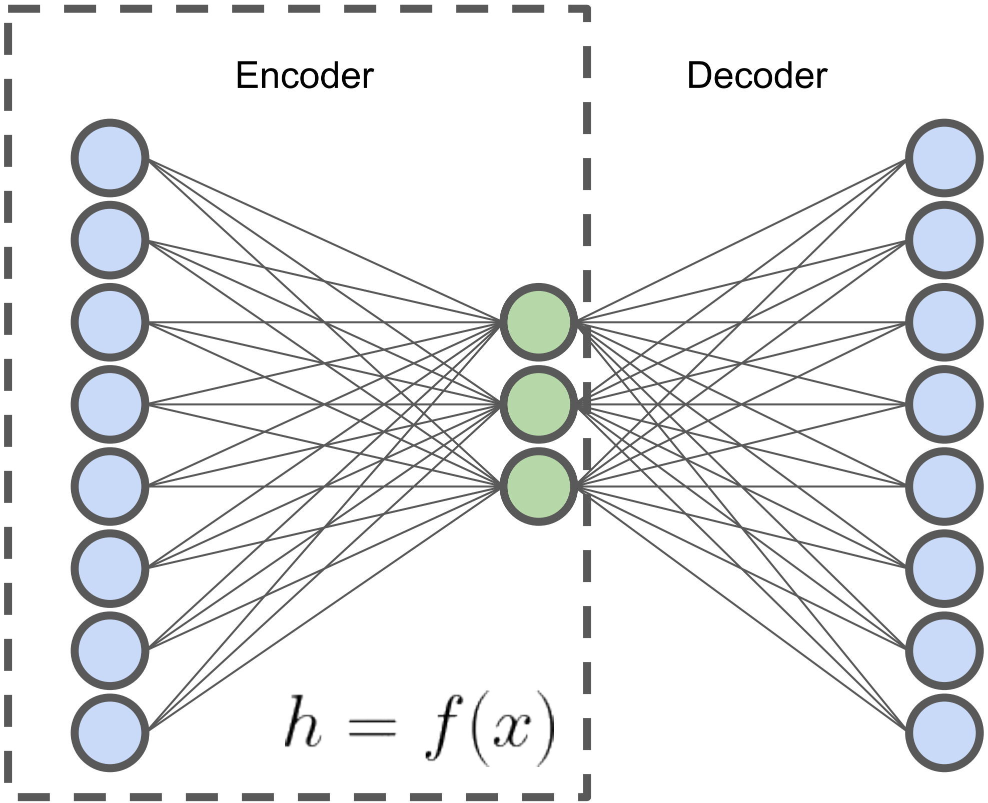
Autoencoder Architecture



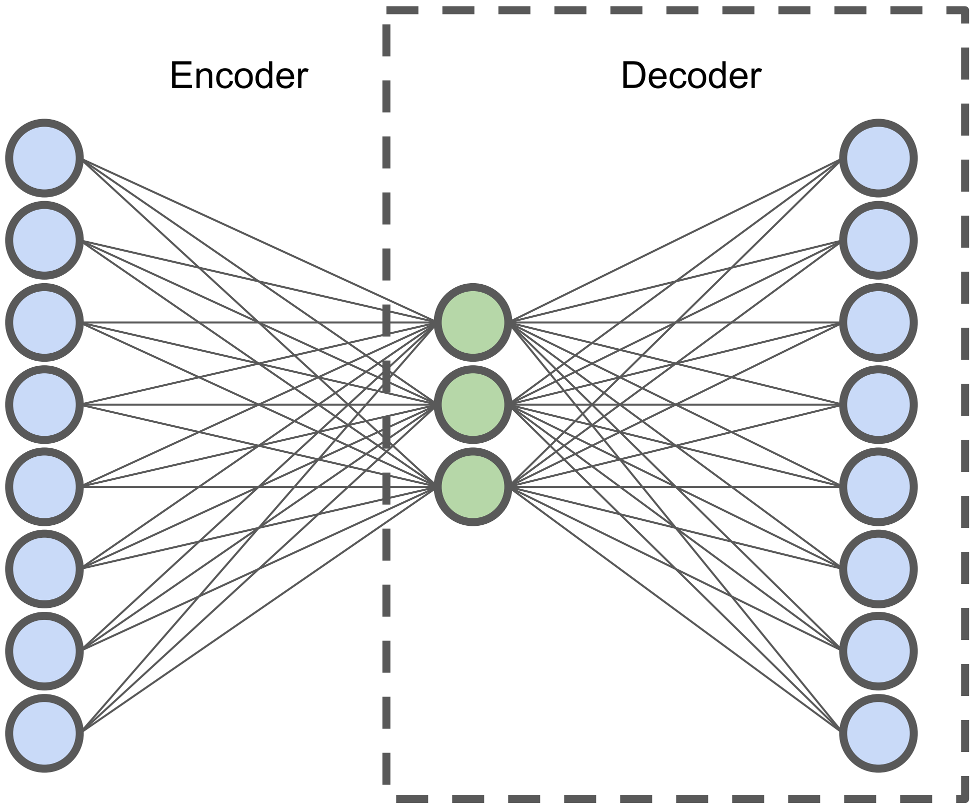
Autoencoder Architecture



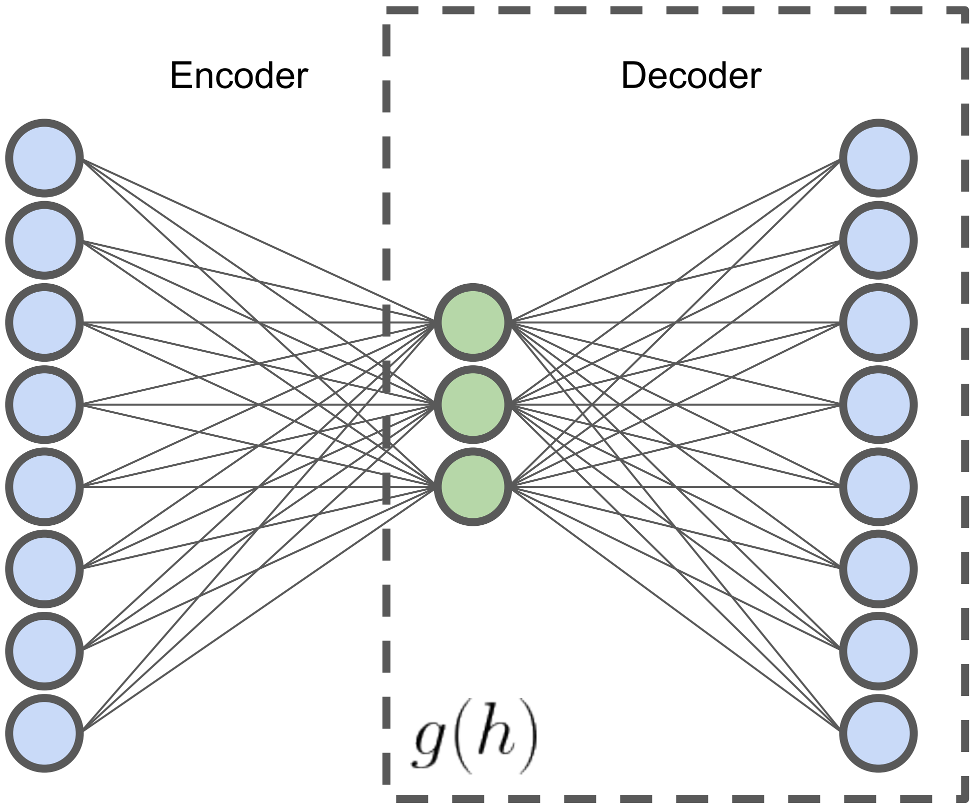
Autoencoder Architecture



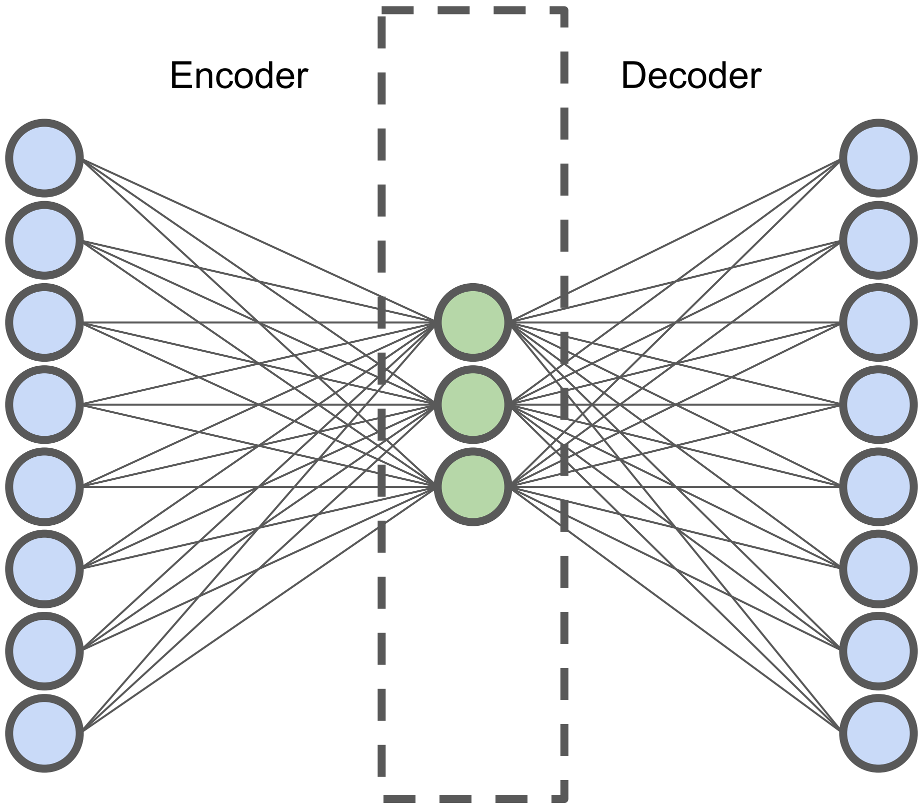
Autoencoder Architecture



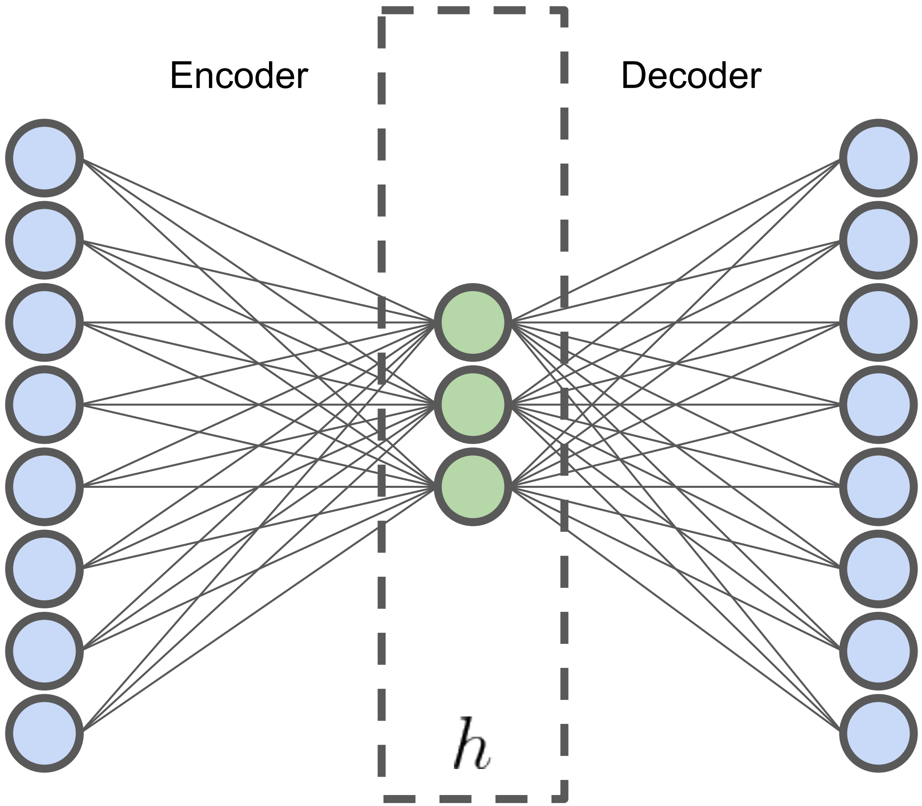
Autoencoder Architecture



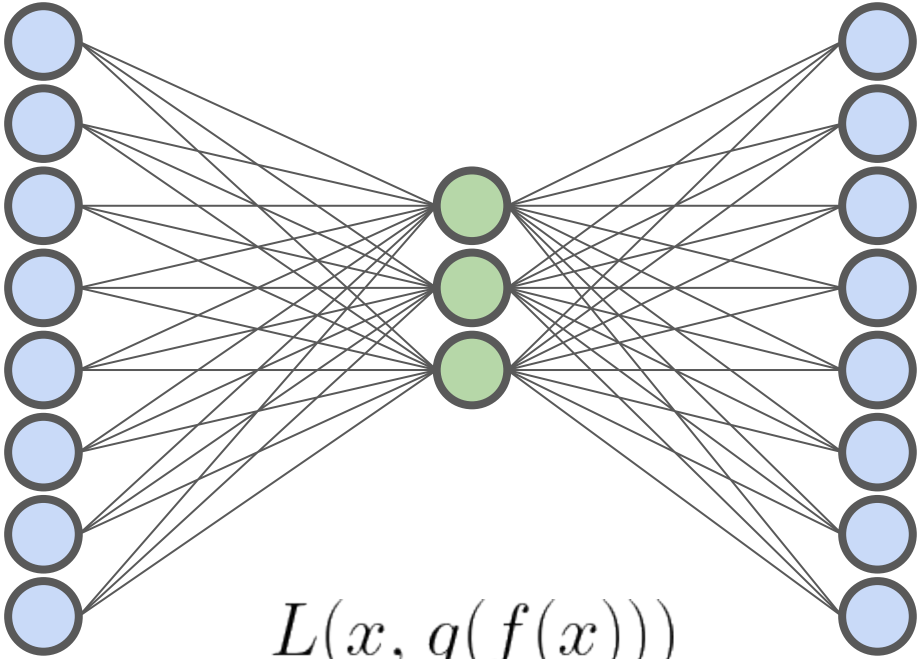
Autoencoder Architecture



Autoencoder Architecture



Autoencoder Architecture

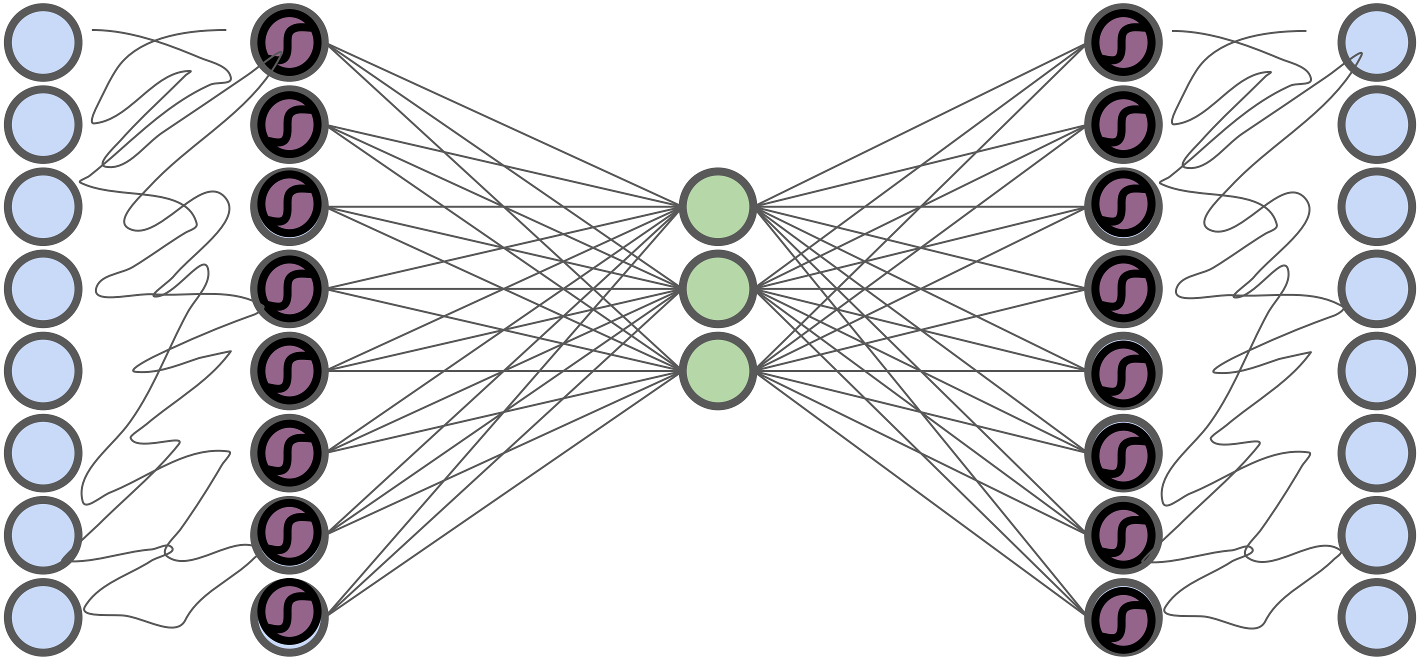


$$L(x, g(f(x)))$$

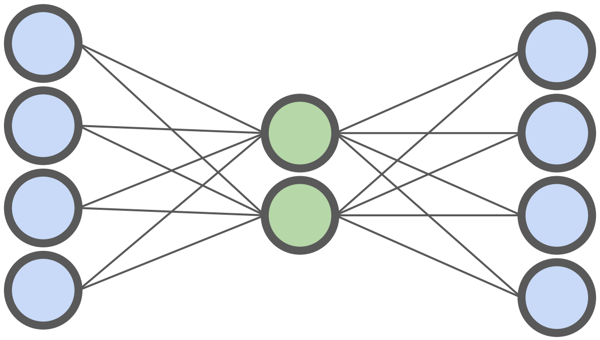
PCA and Autoencoders

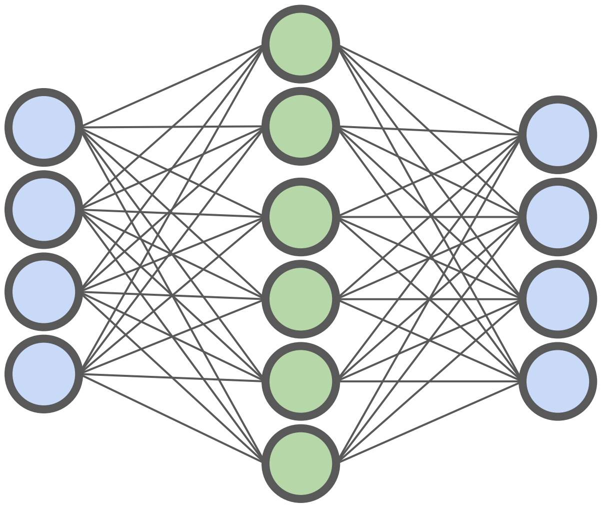
- Encoder is **linear**
- Loss is **MSE**

Adding Nonlinearity

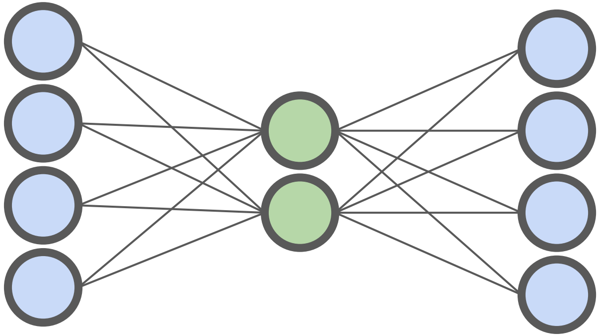


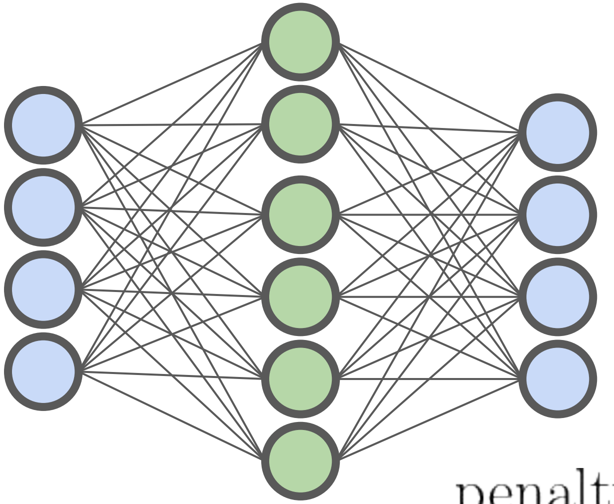
Undercomplete and Overcomplete





Sparse Autoencoders





$$\underbrace{L(x, g(f(x)))}_{\text{loss}} + \overbrace{\Omega(h)}^{\text{penalty}}$$

Penalizing Derivatives

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

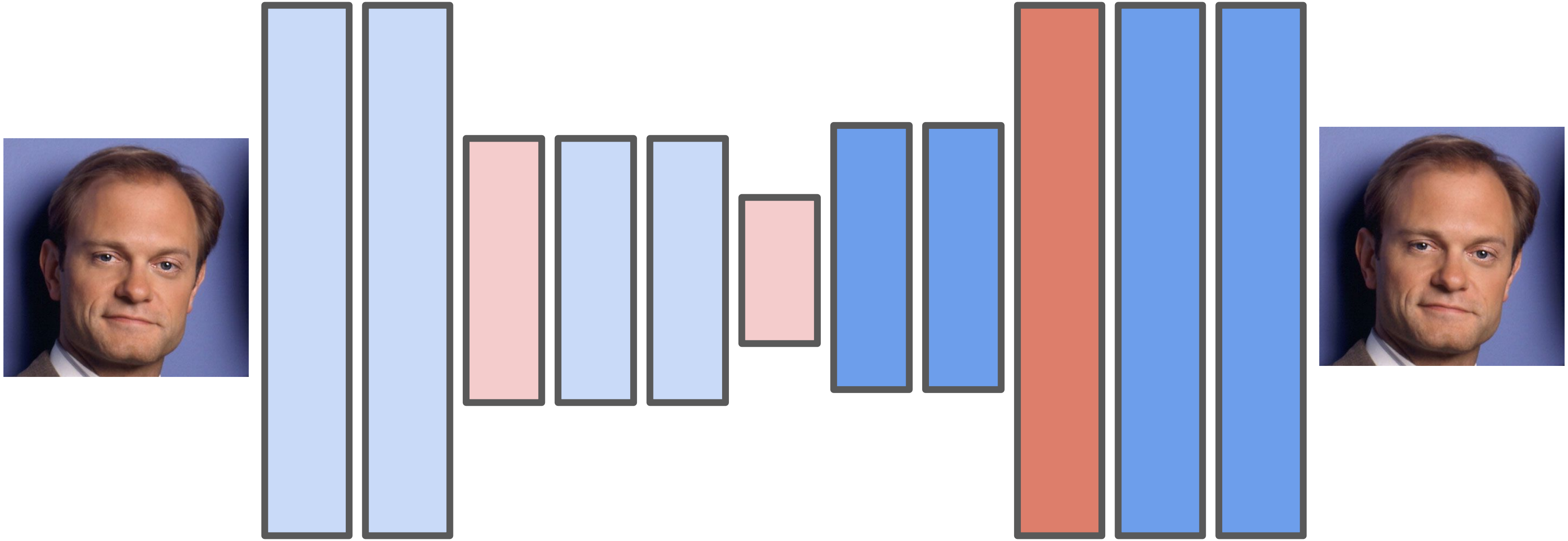
$$\boxed{\Omega(h, x)} = \lambda \sum_i \|\nabla_x h_i\|^2$$

Denoising Autoencoder (DAE)

Add a little noise to x

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

Convolutional Autoencoders



Denoising Convolutional Autoencoder