Deep Learning Architecture [1]
- Shallow architecture
  - Single layer neural network
- Deep architecture
  - Stacked Autoencoder (SDA) [2]

Stacked Autoencoder (SDA)
- Basic Autoencoder. Maps input into hidden, and maps back to reconstruct.
  - Maps input data \( x \) into hidden representation \( h \),
    \[ h = Wx + b \]
  - Maps the hidden representation \( h \) back to reconstruct the original input \( x' \),
    \[ x' = \sigma(W' h + b') \]
  - Learn the parameters \( W, b, \sigma, W', b' \) by minimizing the reconstruction error
    \[ L(W, b, x, x') = \frac{1}{2} \sum_{i=1}^{n} (x_i - x'_i)^2 \]

- Denoising Autoencoder (DA). Corrupt input, and reconstruct uncorrupted input
  - Corrupt the input data by setting \( p \) percent features to zero.
    \[ \tilde{x}_i = \text{corrupt}(x_i, p) \]
  - Maps \( \tilde{x}_i \) to \( \tilde{h}_i \), and then back to \( \tilde{x}'_i \), to reconstruct the original uncorrupted input
    \[ L_{DA}(W, b, \tilde{x}, \tilde{x}') = \frac{1}{2} \sum_{i=1}^{n} (x_i - x_i')^2 \]

- Stacked Denoising Autoencoder (SDA).
  - Feeding hidden representation of \( k^{th} \) DA as input into the \( k+1^{th} \) DA.

Linear Denoiser (LIDE)
- Algorithm inspired by SDA
- Closed-form solution for denoising
- Use infinitely many copies of noisy data to denoise

V\[ W(W^T W)^{-1} P = X \]
\[ Q = (X^T X)^{-1} \]
\[ h = Wx \]

Virtual denoising:
\[ m \rightarrow \infty \]
\[ W = E(W | E(Q)^{-1} \]
\[ Q = \sum_{i=1}^{n} x_i x_i^T \]

The probability of each elements in \( Q \) not being corrupted is:
- off-diagonal, two features
- diagonal, same feature
- last row, one feature, one constant
- last column, one feature, one constant
- the lower right element, two constant

Stacked Denoiser (SLIDE)
- Non-Linearity: Thresholding, "squashing" function between layers, \( T(a) = a_{\geq \sigma} \in (0, 1) \)
- Stack:
  - Feeding hidden representation of \( k^{th} \) LIDE as input into the \( k+1^{th} \) LIDE.
  - \( x^{k+1} = h^k = W^k T(h^{k-1}) \)
- Learn denoising transformation matrix \( W^{k+1} \), by minimizing the reconstruction error:
  \[ W_{opt}^{k+1} = \arg \min_{W} \sum_{i=1}^{n} (x_i - W_i h_i)^2 \]

RBF kernel SVM
- Kernel-trick, Radial Basis Function (RBF) kernel [3]
  \[ k(x, z) = \exp \left( -\frac{\|x - z\|^2}{\sigma^2} \right) \]

References