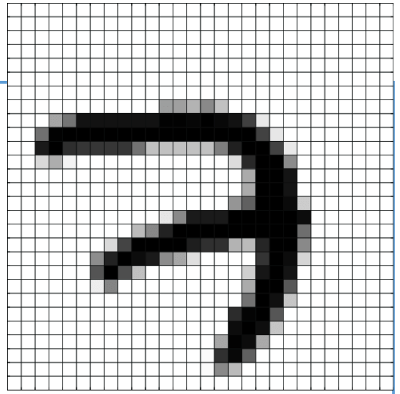
# Principle Component Analysis

# Dimensionality Reduction

## True vs Observed dimensions

## Curse of dimensionality

- Real datasets have high dimensions
- ML methods are statistical
- Density of observations reduces as dimensionality increases



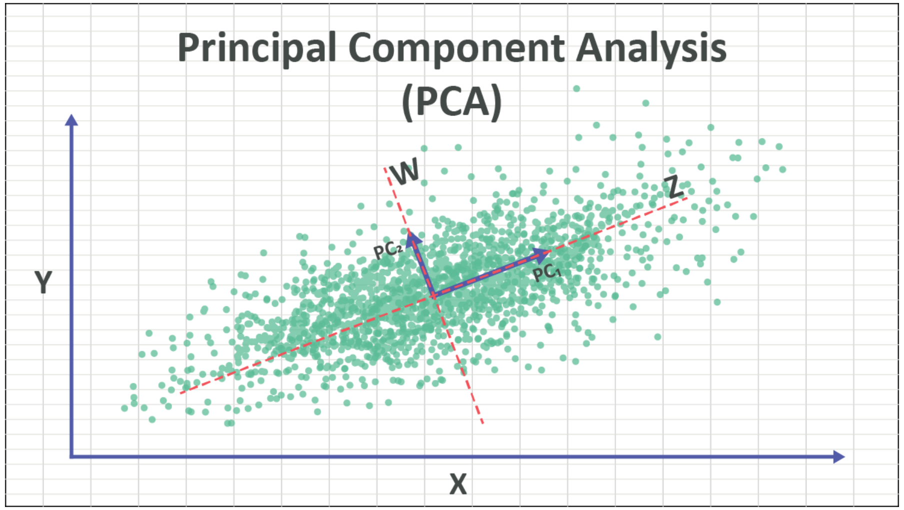
28 x 28

# Dimensionality Reduction

- Can be solved with methods specific to domains
  - SIFT, SURF, HOG, etc.
- Solved based on assumptions of the attributes
  - Independence
  - Symmetry
- Reduce the dimensions of the data
  - Create a new set of dimensions

# Principal Component Analysis

- Defines a set of principal components
  - First direction – with highest variance
  - Second direction – with second highest variance (perpendicular) and so on..
- Take first m components, m<<d

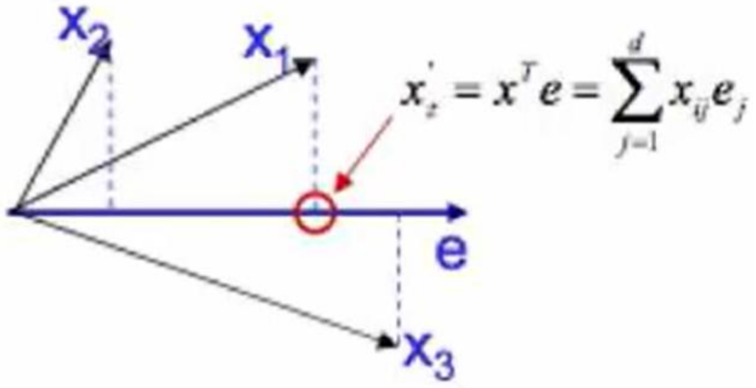


# Principal Component Analysis - EVD

- Center the data
  - $X_i = X_i - \mu$ (subtract mean)
- Compute Covariance Matrix $\Sigma$
  - Covariance of dimensions $x_1$ and $x_2$
- Multiply a vector with $\Sigma$ :
  - Turns towards the direction on variance
- Find vectors **e** which do not turn anymore
  - Eigenvectors
  - Corresponding eigenvalues

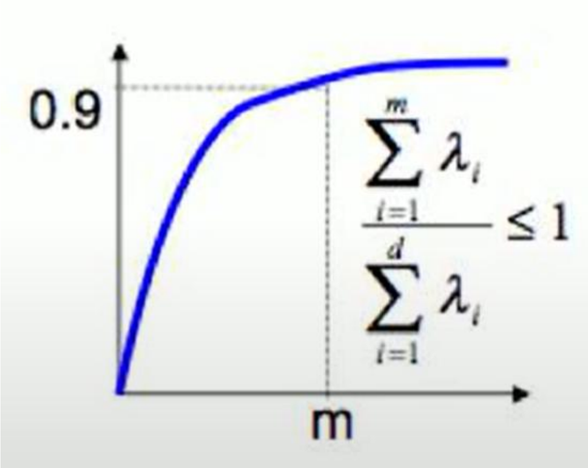
# Principal Component Analysis

- Direction of greatest variability -> eigenvector
- Points $\mathbf{x_i}$ projected onto vector $\mathbf{e}$



# Principal Component Analysis

- Eigenvalue = variance along $\mathbf{e_i}$
- From $\mathbf{e_1} \ldots \mathbf{e_d}$, choose $\mathbf{e_m}$ s.t. m << d
- Where to stop choosing?
  - Sort eigenvectors s.t. $\lambda_1 >= \lambda_2 >= \lambda_3 \ldots >= \lambda_d$
  - First m that explains 90-95% of the total variance.



# Steps to solve for PCA

- Compute mean of the given samples for respective features
- Compute Covariance Matrix

$$\mathrm{cov}(\mathrm{x}_1, x_2) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

  - The value of covariance can be positive, negative or zero.
- Compute Eigenvalues and Eigenvectors
  - Let A be a nxn matrix and X be a non-zero vector for which:

$$AX = \lambda X$$ for some scalar value $\lambda$

# Steps to solve for PCA

- The equation can be written as:

$$AX - \lambda X = 0$$
$$(A - \lambda I)X = 0$$

  - Where I corresponds to the identity matrix (nxn).
  - The condition holds true if (A - $\lambda$I) is non-invertible.
  - $|A - \lambda I| = 0$
  - Find values of $\lambda$ and proceed to calculate the respective eigenvectors.
  - Analyse the $1^{st}$ and $2^{nd}$ (and so on) principal eigenvectors.