Lecture 9-10
Data Compression Based Methods

Ref: Outlier Analysis, Charu C Agrawal
Ref: Tutorial - http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

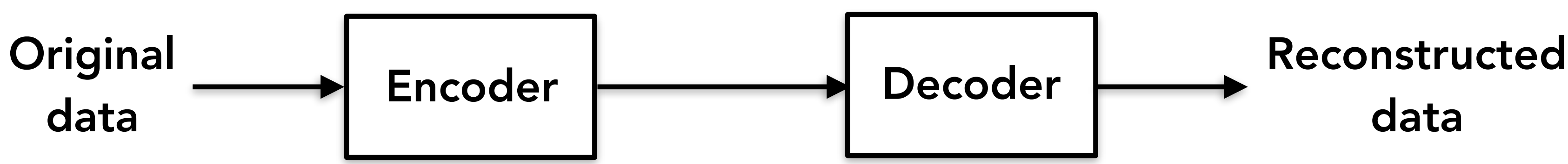
Warmup- Covariance Vs Correlation

$$Cov(X,Y) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{(}x))}{N} \qquad Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_x \ \sigma_y}$$

- Both the terms capture relationship between two variables of a dataset, how change in one variable is related to change in another
- Covariance captures the relationship (variability) in actual units, while correlation captures relationship strength on a standardised scale (-1 to +1)
- We particularly use correlation when we have to compare relationship across datasets

Data Compression

- Represent the same data in a fewer number of bits



Types

- Lossy compression: causes permanent loss of information, the reconstructed data is close to original, but never the same
- Lossless compression: the reconstructed data is exactly same as original data
- Examples?

Main Idea (Lossy Compression)

- Train the encoder/decoder to compress the normal data.
- When used on normal data, we will get low reconstruction error.
- When used on the abnormal data, the reconstruction error will be large.

Main Idea (Lossless Compression)

- Train the encoder/decoder to compress the normal data.
- When used on normal data, we will get compressed easily.
- When used on the abnormal data, the data will not be compressed well.

Principle Component Analysis

- Represent d dimensional data with k dimensions
- Approximate the original d dimensional data with k dimensional data
- The error in approximation is anomaly score

Properties of the new Axis system

- The dimensions are orthogonal to each other
- Hence, the variance of the transformed data is equal to the Eigen vector and covariance is zero
- In other words, there is no correlation left among the axis

Projection to New Axis System

- The first step of PCA is to project the data to a new axis system
- The Eigen vectors provide the new axis system with orthonormal dimensions
- For lossless representation, the number of axis in the new representation system is the same as old one

Steps

1. Mean-centre the data
2. Calculate d*d covariance matrix
3. Obtain Eigen values and Eigen vectors of the covariance matrix
4. Sort the Eigen vectors in decreasing order of Eigen value and keep top k ($k<d$) in a matrix W
5. Project data to the lower dimension, $D_p' = D_p W$
6. Reconstruct the original data with $D_p'' = D_p' W^{-1}$(if $k<d$, W^{-1} won't be a square matrix, use zero padding to get inverse)
7. Take the reconstruction loss as anomaly score $L = |D_p - D_p''|$

Observation

- The variances of data points along the Eigen vectors with low Eigen values are low (variance = Eigen value)
- If a data point deviates too much from the mean value along this direction, it may be an anomaly

Another Anomaly Score

1. Take the d-k Eigen vectors corresponding to d-k lowest values
2. Take the data point, project along these Eigen vectors
3. Calculate the squared sum of these projections normalised by corresponding Eigen value

$$S = \sum_{j=1}^{k} \frac{p_j^2}{\lambda_j}$$

PCA Vs Linear Regression

- In PCA, k can be anything, linear regression restricts k to d-1
- In Linear regression the d-1 dimensions are original dimensions whereas in PCA these are new orthonormal dimensions obtained through PCA

PCA Example

$$S = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}$$

2D data with 3 data points

Eigen Vector

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 24 \\ 16 \end{pmatrix} = 4 \times \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

- In the second multiplication, the resulting vector is just a scaled version of the multiplying vector, such vectors are called Eigen vectors
- A scaled version of Eigen vector is again an Eigen vector, therefore, most libraries give unit vectors (with unit magnitude)

Eigen Values

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 24 \\ 16 \end{pmatrix} = 4 \times \begin{pmatrix} 6 \\ 4 \end{pmatrix}$$

- We see that even in both cases, the transformation of the Eigen vector is by the same amount
- This scaling factor is property of the Eigen vector, and it is called Eigen value

Finding Eigen Values

$$C = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$$Cv = \lambda v \Rightarrow |C - \lambda I| = 0$$

$$\begin{vmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 - 4\lambda + 3 = 0 \Rightarrow \lambda_1 = 3, \text{and } \lambda_2 = 1$$

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Axis Transformation

$$P = SU = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ -1 & -1 \end{bmatrix}.$$