Principal Component Analysis for Dimensionality Reduction

Consider a d-dimensional data set ~~containing~~ containing n records.

Firstly, standardize the given data set.

Now, compute the covariance matrix C or Σ where,

$$\Sigma \in \mathbb{R}^{d \times d}$$

The covariance between any two features x_i & x_j is given by,

$$cov(x_i, x_j) = \sigma_{ij} = \frac{1}{n} \sum_{k=1}^{k=n} (x_i^{(k)} - \mu_i)(x_j^{(k)} - \mu_j)$$

Here, μ_i & μ_j are sample means of features i and j respectively.

For the given data set Σ will be

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\ \vdots & \sigma_2^2 & & \\ & & & \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_d^2 \end{bmatrix}$$

Scanned with CamScanner

Next obtain the eigen values and eigen vectors for the corresponding covariance matrix.

Suppose an eigen vector v which satisfies the following condition

$$\Sigma v = \lambda v$$
$$\Sigma v = \lambda I v$$
$$\Sigma v - \lambda I v = 0$$
$$(\Sigma - \lambda I) v = 0 \quad -①$$

This is similar to homogenous system of linear equations

For $\vec{v} \neq \vec{0}$ and eqn ① to be true

$$|\Sigma - \lambda I| = 0$$

Solving the determinant we obtain d values of λ which represent eigen values corresponding to d eigen values we get d eigen vectors by putting value of λ in equation 1 and solving for v

Now, out of d eigen vectors select K eigen vectors corresponding to K highest eigen values

Now, take those K eigen vectors in descending order of their corresponding eigen values and construct a projection matrix W.

$$W \in R^{d \times k}$$

The K columns of W represent principal components

Using W we can transform a sample vector x, onto the PCA subspace obtaining x'.

$$\underset{1 \times k}{x'} = \underset{1 \times d}{x} \, \underset{d \times k}{W}$$

Similarly, we can transform the entire data set onto K principal components by calculating the matrix product

$$\underset{n \times k}{X'} = \underset{n \times d}{X} \, \underset{d \times k}{W}$$

The data set reduced to K features. Thus our goal of dimensionality reduction is achieved.