Module 4

Optimal Projection parameters

In the last session we setup the PCA objective, and in this session we will determine our first set of ~~obj~~ optimal parameters.

We make 2 general assumptions, in the beginning

(1) First, we have centred data "that means the expected value of our data set is 0,"

$$E[X] = 0$$

Second, is that "the basis vectors form an orthonormal basis"

$$b_1, \dots, b_M \quad \text{ONB}$$

For previous session we carry over the following result:

(A) First we can write our projected data point $\tilde{X}_n$ as a linear combination $\beta_{j,n}$ times b_j, where ? b_j or ? $b_{j,n}$ form the orthonormal basis of our subspace

$$\tilde{X}_n = \sum_{j=1}^{M} \beta_{jn} b_j \quad (A)$$

(B) Our loss function is the average squared reconstruction error between our original data point projection

$$J = \frac{1}{N} \sum_{n=1}^{N} \| X_n - \tilde{X}_n \|^2$$

(C) The partial derivative of our loss function w.r.t. $\tilde{X}_n$ is given by this expression

$$\frac{\partial J}{\partial \tilde{X}_n} = -\frac{2}{N} (X_n - \tilde{X}_n)^T$$

And now we are ready to compute the partial derivative of J wrt β_{in} parameter as follows:

"∂J w.r.t $\partial \beta_{in}$ is the ∂J w.r.t $\partial \hat{x}_n$ times $\partial \hat{x}_n$ w.r.t $\partial \beta_{in}$"

$$\frac{\partial J}{\partial \beta_{in}} = \frac{\partial J}{\partial \hat{x}_n} \boxed{\frac{\partial \hat{x}_n}{\partial \beta_{in}}}$$

So now we going to have a closer look at above

" $\frac{\partial \hat{x}_n}{\partial \beta_{in}}$ is simply given by b_i for i equal 1, … M"

$$\frac{\partial \hat{x}_n}{\partial \beta_{in}} = b_i \quad i = 1, \ldots, M$$

B

And reason for this if we take the derivative w.r.t to one fixed to β_{in}, then only the ith component of the sum (A)(page 2) will play a role, that's the reason why we end up simply to with b_i B

4

That also means that our derivative of ∂J w.r.t.

$\partial \beta_{in}$ " ~~from given by the~~ -2 over N times "

$\hat{x}_n$ means $\hat{x}_n$ transpose b_i, ~~the~~

© paste equation ©

$$\frac{\partial J}{\partial \beta_{in}} = \underbrace{\frac{-2}{N}\left(x_n - \hat{x}_n\right)}^T b_i$$

added b_i

What we going to do now, we going to replace $\hat{x}_n$ using equation (A)

$$= \frac{-2}{N}\left(x_n - \sum_{j=1}^{M} \beta_{jn} b_j\right)^T b_i$$

$$\overset{ONB}{=} \frac{-2}{N}\left(x_n^T b_i - \beta_{in} b_i^T b_i\right)$$

where we exploited that the b_i form an orthonormal basis.

If we multiply b_i to both component here, we end up with sum of β_{jn} times b_j transpose b_i, and since $\beta_{jn} b_j^T x b_i$ is 1, if and only if, $i = j$, otherwise zero (0).

we end up "$b_i^T b_i = 1$"

$$\therefore = \frac{-2}{N}\left(x_n^T b_i - \beta_{in}\right)$$

$= 0$ we need to set this to zero in order to find β_{in} parameters.

"this is zero, if and only if, β_{in} parameters are given by $x_n^T b_i$"

$$\Longleftrightarrow \quad \beta_{in} = x_n^T b_i \quad \underline{\text{Equation (D)}}$$

what this means is that the optimal coordinates of $\hat{x}_n$ w.r.t our basis, are the orthogonal projections of the coordinates of our original data points onto the ith basis vector that spans our principal subspace.

In this session we determined that the coordinates of lower dimensional data is the orthogonal projection of the original data onto the basis vectors that span the principal subspace

In next session we will determine the orthonormal basis that spans the principal subspace.