Module 4

Reformulation of the objective function.

In the last session we determined the coordinates of the optimum projection w.r.t the orthonormal basis that spans our principal subspace

Before we go on to determine the optimum/mal basis vectors, lets rephrase our loss function First.

The result that we have so far:

(A) Description of our projected datapoint:

$$\hat{x}_n = \sum_{j=1}^{M} \beta_{jn} b_j$$

(B) Our Loss function:

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \hat{x}_n \|^2$$

(c) The partial derivative of our Loss function w.r.t. our projected data point

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

(1) optimal coordinates that we found in last session.

$$\beta_{j,n} = x_n^T b_j, \quad j = 1, \ldots, M.$$

Before we go on, and determine the optimal basis vectors, Lets rephrase our loss function. This will make it much easier to find our basis vectors.

For this, Lets have a closer look at the difference vector, between our original data point and our projected data point

"We can write $\hat{X}_n$ is given by equation A, which is the sum $j=1$ to M, β_{jn} times b_j"

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

If we now use the results of our optimal β_{jn} parameters, (D)
we get" sum $j=1$ to M, X_n transpose times b_j times b_j" use (D)

$$\overset{(D)}{=} \sum_{j=1}^{M} (X_n^T b_j) b_j$$

Now we rewrite this in the following way:

"$(X_n^T b_j)$ is just a scalar or dot product in this particular case, and dot products are symmetric, so we can swap the order, and we can also move the scalar to the end" ↘ see above

So what we end up with: (projection matrix) (1)

$$= \sum_{j=1}^{M} b_j (b_j^T x_n) = \left(\sum_{j=1}^{M} b_j b_j^T \right) x_n$$

(move x_n out of sum)

and if we look at this (1), it's a projection matrix, this means that $\hat{x}_n$ is the orthogonal projection of x_n onto the subspace spanned by the M basis vectors b_j where $j = 1, \ldots, M$

$$\left(\sum_{j=1}^{M} b_j b_j^T \right)$$

Similarly we can write x_n as (Sum) Sum $j=1$ to M of b_j times b_j transpose times x_n

$$x_n = \left(\sum_{j=1}^{M} b_j b_j^T \right) x_n \quad \leftarrow (8)$$

$$+ \left(\sum_{j=M+1}^{D} b_j b_j^T \right) x_n \quad \leftarrow (9)$$

(x) we write x_n as a projection onto the principal subspace plus (y) a projection onto the orthogonal complement

(Remember $\tilde{x}_n$ is the approximation onto x_n)

So if we now look at the difference vector between $\tilde{x}_n$ and x_n, what remains is exactly this term (y) page 4

$$\therefore \quad x_n - \tilde{x}_n = \left(\sum_{j=M+1}^{D} b_j b_j^T \right) x_n$$

(x)

So now we can look at this displacement vector, the difference between x_n and its projection ($\tilde{x}_n$), and we can see that the displacement vector, is exclusively in the subspace we ignore, i.e. (x)

that means, the orthogonal complement to the principal subspace.

Lets look at an example in 2Dim. (see pic)

we have a data set in two Dim, represented by those dots, and now we interested in projecting them onto the U_1 subspace.

when we do this and look at the difference vector between the original ~~vector~~ data and projected data, we get these vertical lines, that means they have no x component, no variation in x, that means they only have a component that lives in subspace U_2, which is the

orthogonal Complement to U_1, which is the subspace that we projected onto (see pic)

So with this illustration, let's quickly rewrite this in a slightly different way:

$$= \sum_{j=M+1}^{D} (b_j^T x_n) b_j \quad \text{equation (E)}$$

We looked at the displacement vector between x_n and its orthogonal projection onto the principal subspace $\tilde{x}_n$

And now we going to use this to reformulate our loss function

So from equation (B) (8)

" we get that our loss function is 1 over N times Sum n=1 to N, of $x_n - \hat{x}_n$ squared, average squared reconstruction error. "

$$J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \hat{x}_n \|^2$$

(r)

And now we going to use equation (E) for the displacement vector here (r)

We rewrite this now, using equation (E):

" as 1 of N times Sum n=1 to N, and now we going have inside that Squared norm, the expression here (E); sum j=m+1 to D of b_j transpose by Transpose times x_n times b_j squared. "

$$\overset{(E)}{=} \frac{1}{N} \sum_{n=1}^{N} \left\| \sum_{j=M+1}^{D} (b_j^T x_n) b_j \right\|^2$$

And now we going to use the fact that the b_j form an orthonormal basis which will greatly simply the expression:

$$\underset{\text{ONB}}{=} \frac{1}{N} \sum_{n=1}^{N} \sum_{j=m+1}^{D} \left(b_j^T x_n\right)^2$$

identical.

and now we going to multiply it out explicitly

$$= \frac{1}{N} \sum_{n} \sum_{j} b_j^T x_n x_n^T b_j$$

Now we rearrange the sums, we going to move the sum over j outside:

$$= \sum_{j=m+1}^{D} b_j^T \left(\frac{1}{N} \sum_{n=1}^{N} x_n x_n^T \right) b_j$$

S

Now if we look very carefully we can identify this expression (above)

& the data covariance matrix S, because we assumed we have centered the data.

Σ mean of data is 0

This means now we can rewrite our loss function using the data covariance matrix and we get: , our loss is the ... sum ...

$$= \sum_{j=m+1}^{n} b_j^T S b_j$$

we can also use a slightly different interpretation.

by rearranging a few terms and using the trace operator

$$= \text{trace}\left(\underbrace{\left(\sum_{j=m+1}^{n} b_j b_j^T\right) S}_{\text{projection}}\right)$$

Can also interpret this matrix as a projection matrix

the projection matrix, take our data covariance matrix and project it onto the orthogonal complement of the principal subspace.

That means we can reformulate our loss function as the variance of the data projected onto the d subspace that we ignore

Therefore minimizing this loss, is equivalent to minimizing the variance of the data that lies in the subspace that is orthogonal to the principal subspace

In other words, we interested in retaining as much variance after projection as possible.

The reformulation of the average squared reconstruction error i.t.o. data covariance gives us an easy way to find a basis vector of the principal subspace, which we will do in next lesson: