

SLIDE 1: The multi subjects analysis are very important in the Neuroscience field because permits to compare studies across subjects, and generalize the discoveries founded in the single subject analysis. However, the anatomical and functional structure of the brain is different between subjects even if we have identical synchronized stimuli. For that, an alignment step is necessary in order to improve the analysis. The first type of alignment that we can find in literature is the anatomical alignment proposed by Telaric and turnu. The images are aligned to a common template, telaric template, using affine transformation. However, it fails to align the functional structure of the brains. For that reason, Haxby proposed a functional alignment called hyperalignment, that uses procrustes transformation.

SLIDE 2: In order to understand the procrustes transformations in the case of fMRI data, I will explain how we can express these data as matrices. So having for example a subject watching images of faces and objects during the scans, the level of activations for each voxel and for each time point, that correspond to a specific stimulus, is represented by a matrix, where the rows are the representation of the activation of the brain having this specific stimulus, as for example here the first row is time one and correspond to the activation when the subject watch a face, the second row the subject watch a shoes and in the third a chair. So the first element represent the activation respect to the face stimuli for the first voxel. In two dimensions, so having only two voxels we can represent each stimuli in a space. So, in general we have a space with dimension equals to the number of voxels.

SLIDE 3: So, having N subjects and a matrix X_i with dimension n times v , now we can see that the alignment problem is referred to the column of our matrices. The rows are the response stimuli activation, the stimuli are time synchronized, however the columns that represent the time series of activation of a voxel could be not in correspondence between subjects. To resolve this, we recall the Procrustes method that uses orthogonal matrices in order to match a matrix into a target one as close as possible. The distance is computed by the frobenius norm. So, the orthogonal procrustes problem is expressed in this way: find the orthogonal matrix R that minimize the distance between one subject matrix onto another subject matrix.

SLIDE 4: The method proposed by Haxby, called hyperalignment, is a sequential application of the procrustes method, so we can not propose a statistical approach or a clear optimization criteria.

For these reasons, we re phrase the procrustes problem in terms of a statistical model. So having our matrices X_i with dimension n and v , we can define it in these way,

SLIDE 5: where ϵ is the error term having a multivariate normal matrix and M is the mean matrix. So we align each X_i to the common mean, the distance between it is the error term that we want to minimize. Schonemann, sorry for the pronunciation, found that the minimum is reached when $R_i = UV^T$ where U and V comes from the singular values decomposition of $X_i^T M$. We will find that it s equals to the maximum likelihood estimation.

SLIDE 6: So, the our idea is to use a prior distribution for the rotation matrix in order to put some obviously prior information and then analyze the most plausible (plausible) rotation. The matrix R_i is an orthogonal matrix, so the prior distribution must takes values in a Stiefel manifold. The distribution that we consider is the matrix fisher von mises expressed in this way, where k is the concentration parameter and Q is the location parameter.

SLIDE 7: So, we can express the likelihood and the log likelihood in this way, assuming that voxels have independent errors with equal variance.

SLIDE 8: and the maximization problem is expressed as follow, that is essentially equals to the procrustes formulation presented previously. After some computation, we find the same solution for R_i , but we decompose this quantity instead of this. So, in the maximum likelihood estimation some information about the matrix rotation enters in the singular value decomposition step.

SLIDE 9: in the multi subjects case, we have simply the joint likelihood because the subjects are independent.

SLIDE 10 We don't have a closed solution, we propose so the same iteration algorithm of the Generalized Procrustes solution, but we decompose this quantity instead of only this.

SLIDE 11: This is the Generalized procrustes algorithm. In short, we start with a reference matrix M , and we align all matrices into the space of this matrix, and so we recompute the mean considering the new transformed matrices, and we compute the distance between the old and the new matrix mean. The algorithm finish when this distance is less than some value imposed by the user or if we did more iteration than the number of iteration imposed by the user.

SLIDE 12: So, how we can choose the prior parameters of the matrix fisher von mises distribution? We want to insert prior information about our rotation matrix. We can think that closer voxels have similar rotation loading than voxels that are far. In this way, we can think about Q as a similarity matrix, that takes high values if the voxels are near and small values if the voxels are far. The most simple idea is to construct Q as similarity matrix using the euclidean distance of the coordinate of the voxels. So having the three coordinates for each voxels. However, each type of distance can be used, we used the euclidean one in order to have a simple interpretation, and also because in this case the matrix has full rank, and if Q has full rank, it has a unique polar decomposition, Mardia found that the distribution has mode when R is equals to the polar part of the location parameter. So, in the case of full rank the mode is unique. There are few literature about the case of matrices without full rank.

SLIDE 13: So, we apply this method to real data, faces and object datasets that analyze the ventral temporal cortex of 10 subjects watching some photos of human and animal faces and objects. We align these images using Anatomical Alignment, Hyperalignment, GP and GP with prior and then we use a linear support vector machine in order to classify the brain activation expressed into the voxels, using leave one out subjects crossvalidation, and leave one out runs to estimate the alignment parameters. We find that the accuracy of classification is equal to 0.67 using our method respect to 0.31 using the anatomical one. This method reduces the error of classification by 10% respect to the hyperalignment and by 17% respect to the generalized procrustes.

SLIDE 14: As final step, we want to see the difference of our method respect to other method as hyperalignment. The Hyperalignment is a sequential approach of the Procrustes transformation, so it doesn't reach the global minimum imposed by Generalized Procrustes. We find in fact that after hyperalignment the pairwise difference is higher respect to using GP as alignment and so the brains after hyp are more different.

Also, another thing is that the hyperalignment depends on the order of the subjects that enter in the algorithm. Permuting the order of the subjects, we can find different situations and so different classification accuracy. Here, I permuted for 100 times I think, and we can see how the output can change. In fact, this method doesn't reach the global minimum imposed by the GP.

SLIDE 15: Finally, we can see that the generalized procrustes solution doesn't have a unique solution, because recalling the minimization if we multiplied the matrices by an orthogonal matrices the minimization remains the same. So, every rotations of our final solution is a possible solution. In our method we have also the quantity of the prior fisher distribution, so the solution is unique. This is important because, the solution thanks to the prior distribution maintains an anatomical meaning.

SLIDE 16: So concluding, the method proposed doesn't depend on the order of the subjects as hyperalignment, has a unique solution, reaches a global minimum, improves the classification in particular it distinguish between fine category structure, and thanks to the prior we restrict the range of possible solution for the rotation matrix, using anatomical information as the coordinates of the voxels that hyperalignment or gpa doesn't consider. So, it's simpler to understand the meaning of the rotation matrices.