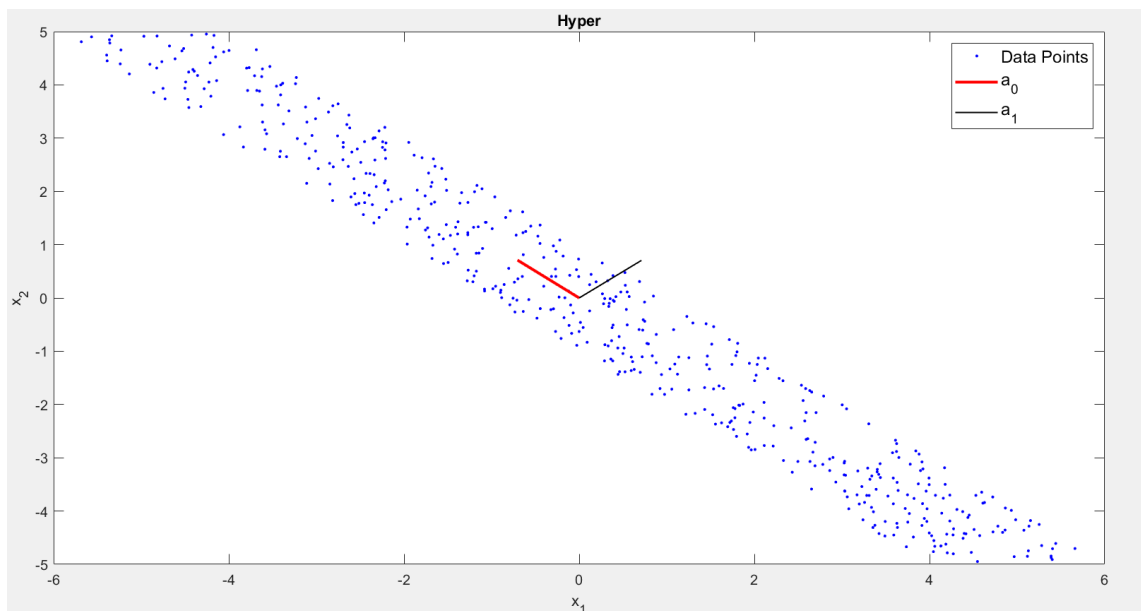
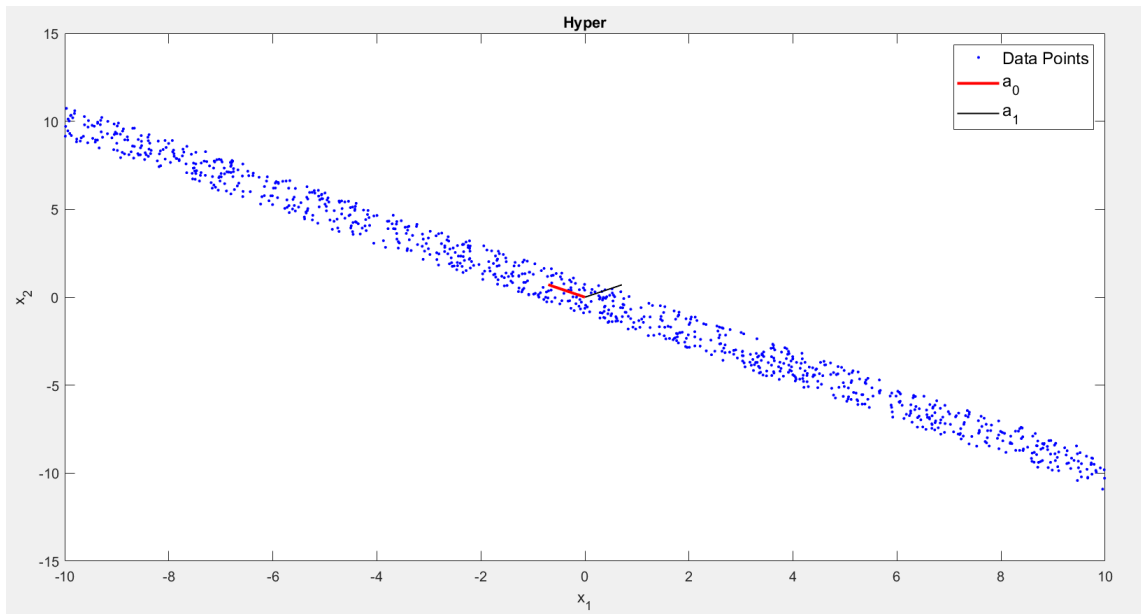


1. The principal components and corresponding covariances are shown in the table below. The component \mathbf{a}_0 has a significantly higher covariance compare to \mathbf{a}_1 and that means the direction of \mathbf{a}_0 is parallel to the line $\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{0}$. The other two figures shown following the table are visual figures of data points and corresponding principle components lines.

	Principal Component	Covariance
\mathbf{a}_0	[-0.7042, 0.7100]	63.5064
\mathbf{a}_1	[0.7100, 0.7042]	0.1656



2. 1) MDL is all about “learning as data compression”. The similarities between MDL, BIC and AIC are they all do model selection method based on information-theoretic ideas. MDL and BIC are often used interchangeably since MDL is derived from BIC. However, they’re different if samples are not i.i.d. and the number of samples is almost infinity. On the other hand, AIC penalizes complex models less and more on model performance on the training dataset and in turn, select more models.

2) To construct an MDL solution for the computer experiment 6.1 from the text, I would figure out how to compress the generated data. The data is generated by using the function “generate_hyper” so the MDL must be applied on that function first. Then I would construct a list of models that will be compared to the data and find out the smallest and best model.

3) Personally, I think the biggest challenge in applying MDL is to find out the best model that fits samples data as well as computing data through the best model. To find out the best model, one must know how the data is originally collected or created so that he or she can give out best options when applying MDL.