3.)

The cross-validation selector is used to select a model that isn’t being overfit by using the test set to calculate the log-likelihood. The value it produces will always be less optimistic than the one computed on the in-sample data, and their average over several test sets will produced a more accurate estimate. One of the issues with the cross-validation approach is that, based on what observation are placed where when the data is divided into training and validation sets, the validation estimate of the test error rate can be highly variable, but this is the kind of thing that the average tries to smooth out. The Bayesian Information Criterion (BIC) method takes an approach of penalizing more complex models, favoring models that are more parsimonious. The BIC selector given a model, will try to maximize the posterior likelihood of the data. For BIC, problems arise when the sample size is not larger than the number of parameters. The Discriminative Information Criterion (DIC) compares the evidence of the model, given the data to the average of the corresponding model to generate data belonging to competing classes (anti-evidence). Comparing the 3 criteria above it appears the DIC method might be more helpful in identifying which model is better. This is because, as mentioned in the paper, “A Model Selection Criterion for Classification: Application to HMM Topology Optimization”, the author goes on about picking the best model isn’t necessarily about Occam’s razor and the simplest model, it’s more about picking the model that’s less likely to misclassify the data, which is what the DIC incorporates with the anti-evidence.

4.)

The best combination turned out to be my custom features paired with the BIC selector with a WER of 0.505618.The custom features generated overall better WER scores than other, as did the BIC selector compared to the other selectors. Most of the custom features involved some kind of distance relation between the left and right hand that the other features, like the normalized and polar coordinate features, were missing altogether, so it might mean that this relation is important in improving the WER score. Also, the 2nd lag difference for image frames was used for left and right hands, so it could be possible that the 2nd lag is of more importance and the single frame lag in classification. It is interesting that the BIC selector generated better WER scores than the other selectors. It seems as though penalizing models more free parameters was more useful than using the anti-evidence. It’s also interesting that the custom feature results using cross-validation and DIC were a tie for 2nd place generating the same results. I’m guessing that both selection methods picked the same final model.

Combining features seems to work well also. Combining my custom features with the polar features generated an “impressive” WER score of 0.426966. Looking into what other features might be helpful, I came across a paper published way back in 2007, titled “Recent developments in visual sign language recognition”. The authors suggest using interesting features pulled by image processing. One of which is using the borders of the hands, clipped from the image, to compute geometric features, such as, the area of the hand, the orientation of main axis (the tilt of each hand), and the ratio of inertia along and perpendicular to the main axis of the hand. I’d love to see how much features like these could improve the WER scores for different selectors.