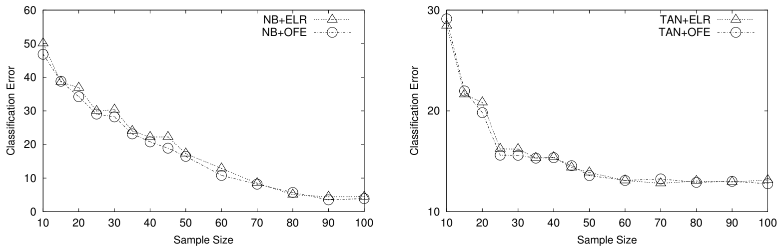

Figure 1: $G > T$ situations, complete data. (a) Model is NB; Truth is $C \equiv E_1$; (b) Model is TAN; Truth is NaïveBayes. (Each point is averaged over 10 runs)

5.4 Model is More Complex than Truth ($G > T$)

Section 5.1 focused on the common situation where G (the BN-structure being instantiated) is presumedly *simpler* than the "truth" — *e.g.*, we used naïve-bayes when there probably were dependencies between the attributes. This section considers the opposite situation, where we allow the model "more degrees of freedom" than the truth. As this is atypical, we could only consider artificial data.

In our first experiment, we attempt to learn the parameters for a naïve-bayes model, when the truth is $C \equiv E_1$ — *i.e.*, the other attributes $E_2, \ldots, E_k$ are each irrelevant. We focus on $k = 6$ and $k = 7$ attributes, where all variables are binary. When the data is complete, we used first OFE and then ELR to instantiate the parameters of a given NaïveBayes model. Figure 1(a) shows the learning curve as we increase the sample size, over 10 different runs. (Each run used its own training sample.) We see that NB+OFE is consistently slightly better than NB+ELR: averaged over all of the runs, this is significant at $p < 0.002$.

We also weakened the $C \equiv E_1$ condition, to simply require C be highly correlated with E_1. Using the same set-up show above, when the correlation is 0.96, we found NB+OFE $\Leftarrow_{(p<0.001)}$ NB+ELR. When the correlation is 0.80, the dominance is even more: NB+OFE $\Leftarrow_{(p<0.0001)}$ NB+ELR.

The second experiment "reverses" the situations shown in Section 5.1.4 (on webpage: [Gre04, #study]). Here, the truth corresponds to a naïve-bayes structure (with no dependencies between the evidence E_i variables, conditioned on the class variable), but we attempt to find the parameters for a "P_m-based structure" — *i.e.*, a TAN structure that links $E_1 \equiv E_2 \equiv \ldots \equiv E_m$. These results appear in Figure 1(b), again this is averaged over 10 runs. (This difference is not significant.)

We next considered the same two situations, but in the *incomplete* data case. In particular, here we blocked a value of any entry with probability 0.2.

The results, shown in Figure 2, show that the generative measures (NB+APN and NB+EM) dominated the discriminative NB+ELR: NB+APN $\Leftarrow_{(p<0.02)}$ NB+ELR and NB+EM $\Leftarrow_{(p<0.015)}$ NB+ELR. (Moreover, NB+EM $\Leftarrow_{(p<0.025)}$ NB+APN.) The generative approach is also superior in the other sitation (Figure 2(b)): TAN+APN $\Leftarrow_{(p<0.025)}$ TAN+ELR, and TAN+EM $\Leftarrow_{(p<0.05)}$ TAN+ELR.

In a nutshell, we observed that discriminative ELR learning typically did worse than the generative learners in this "model is more complex than truth" situation, when dealing with either complete or incomplete data.

References

[Gre04] 2004. http://www.cs.ualberta.ca/∼greiner/ELR.

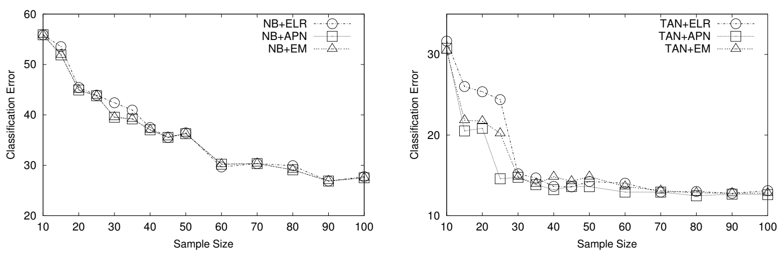

Figure 2: $G > T$ situations, *incomplete* data. (a) Model is NB; Truth is $C \equiv E_1$; (b) Model is TAN; Truth is NaïveBayes. (Each point is averaged over 10 runs.)