Nature vs. *Nurture* (vs. *Nerd*) *

Jason Imbrogno, University of North Alabama

Luis Quintero, Johns Hopkins University

February 27, 2017

*We would like to thank Daniel B. Neill of Carnegie Mellon University for comments.

Abstract

We develop an intuitive model to predict the commission of a crime by juveniles in the US using an extensive set of information on the individuals. We have background data on demographics, family composition, socioeconomic status, family relationships, and schooling outcomes. These variables are broadly categorized into three groups: *Nature* (variables such as family and home environment characteristics which do not directly imply the active involvement of the youth), *Nurture* (variables involving the individuals' relationship with his family that require his active participation), and *Nerd* (variables describing schooling and other studying characteristics such as time spent reading for fun or doing homework). The *Nerd* attributes perform better at predicting crime than *Nurture* and *Nature*. Individually, none of these categories perform well in predicting the crime probabilities. The bundle of individual `Best Predictors`, however, are able to correctly predict crime in 65.5% of the observations. This bundle has attributes from all categories, although *Nerd* is overrepresented. This suggests that it is necessary to address the issue of youth crime in all categories, even if educational policies receive special attention.

1 Introduction

We develop an intuitive model that best predicts the commission of a crime by juveniles in the US using an extensive set of information on the individuals. We have background data on demographics, family composition, socioeconomic status, family relationships, and schooling outcomes. These variables are broadly grouped into three categories: *Nature*, *Nurture*, and *Nerd*. We also know whether the individuals committed any of the following crimes between age 11 and 16: vandalism, petty theft, major theft, motor vehicle theft, property crime, aggravated assault, or selling drugs. The initial goal of this research is to determine which broad category of variables - *Nature*, *Nurture*, or *Nerd* - holds the most predictive power over criminal behavior.

We develop this model in the context of having access to all of this background information for the juveniles involved. But one could easily envision a scenario in which data is more limited, or, even more importantly, where significant costs are involved in obtaining the input data. In that case, knowing the most relevant variables toward crime prediction would help keep the cost of information gathering (for social workers, school district officials, researchers, or any other interested party) to a minimum.

We also want to understand whether there is a gain to knowing details about all three categories instead of just one. Much of our data interweaves and influences each other. For example, demographic characteristics have a lot to do with schooling outcomes, as evidenced by the achievement gap (Coleman et al., 1966) regardless of the outcome variable "crime" that we consider here. It is important to capture these relationships in order to understand as complete a model as possible in predicting crime. It is also relevant to know whether finding information on any single input variable (say family income) yields an information gain on other, perhaps more costly,

input variables (for example, how attached the individual is to his mother, which requires field research and survey completion or interviews). We want to show the interrelationships between all of our attributes that contribute to the prediction of crime. Each category contains many attributes. Those attributes within any given category not only relate to each other, but their data collection processes are similar. For example, most of the attributes under *Nerd* can be obtained directly from school district records for the individuals, while most of the attributes under *Nurture* require field research and surveys or interviews with the juveniles and their families. That is partly the justification for splitting the attributes into categories. However, it is also important to realize that attributes across different categories may be very influential in predicting crime commission. In fact, a subset of attributes that span multiple categories may be better at predicting crime than all of the variables within any given category. We find the set of individual best predictors of crime in this data set and model their predictive power on the commission of a crime.

Overall this project has four explicit goals: understanding which category of variables are most capable of predicting criminal behavior in youth; understanding of whether there is a gain to knowing every category instead of just one; overlap and interweaving of attributes that contribute to the prediction of crime; and individual best predictors of criminal behavior.

Under a situation of limited resources, it is important to prioritize policies related to reducing youth criminal behavior. Each of our goals can help guide policymakers when deciding where to spend public money in the interest of preventing crime. Specifically, it is important to know which possible policy approaches can affect the variables that have a stronger impact on the probability of a youth committing a crime. This is an additional advantage to grouping the attributes by category. If, for example, *Nerd* holds the greatest predictive power over crime, then it makes sense

to spend more public funds on schools and tutoring programs than on *Nurture*-esque parenting classes in the interest of reducing juvenile crime. The results of this project can help direct scarce resources toward the best methods of reducing youth crime by revealing whether policies aimed toward schools, home environment, or family background are the most relevant determinants of criminal behavior.

It is also crucial to determine if any of our attributes can adequately predict criminality. If none of them matter, it reveals one of two possibilities. Either we, as policymakers, have little influence over setting up the proper conditions to keep children from committing crimes or we simply have no idea what variables matter toward criminal behavior. In the latter case, we need to think further outside the box and collect information on other attributes to see if something else can predict. In very broad terms, our imperative policy questions are whether or not juvenile criminal behavior can be reliably predicted with available attributes, and, if it can, which individual attributes and which category of attributes are the best predictors of that behavior.

2 Literature

The debate of *Nature* vs. *Nurture* in determining peoples' life outcomes has been around sociology and psychology circles since the mid-1800s, with Galton (1895) publishing the first work exploring that dichotomy in a formal way. The commission of a crime, especially a serious one, has a negative long-term impact on many aspects of an individual's life, including their freedom, prosperity, and relationships (Sampson and Laub, 1990). For that reason, researchers have looked at criminal behavior as a particularly relevant outcome measure. A plethora of papers has shown that crime is concentrated in minority males from high poverty neighborhoods (Freeman,

1999; Raphael and Sills, 2006). These attributes (gender and income level) in our categorization are considered part of *Nature* and *Nurture*, respectively. Freeman proposes a more theoretical framework for this observation on criminal behavior. Likewise, many papers have explored the so-called achievement gap (for example, see Hochschild, 2003), whereby poorer minorities achieve lower standardized test scores, graduation rates, and college attendance rates than their otherwise similar but affluent, majority counterparts. More recently, some papers have attempted to link student performance or attendance in school to later criminal behavior (Deming, 2011; DeAngelis and Wolf, 2016). We classify these performance variables in a new category called *Nerd*. The aforementioned work does not include the three categories we analyze, and overall uses fewer indicators than this paper, usually limited to basic demographics, some socioeconomic indicators, and limited schooling data such as test scores. After testing the models that use attributes of each category individually, we pool all attributes and select those that improve significantly the predictive power of the model. We call these the group of `Best Predictors`.

3 Data

A description of all the data can be seen in Tables 1 and 2. The tables list the variable name, description, possible values it can have, and the category we include it in, be it *Nature* (A), *Nurture* (U), or *Nerd* (E). *Nature* variables include demographics, residential characteristics, family composition, and economic background. We define *Nature* variables as family and home environment characteristics which do not directly imply the active involvement of the youth. The youth himself is not responsible for or related to the values of these variables except for the fact that he exists. For example, the number of people in the household clearly depends on the youth's existence and

residence at home, but otherwise nothing about the youth affects the value of this variable. *Nurture* variables involve the individuals' relationship with his family and require his active participation. Attributes categorized under *Nurture* include family management variables (for example, how attached the youth is to his mother and how well he gets along with his parents) and variables related to religious and other everyday activities (such as eating dinner with the family). Finally, *Nerd* describes schooling and other studying characteristics such as time spent reading for fun or doing homework. There are 28 *Nature* attributes, 15 *Nurture* attributes, and 12 *Nerd* attributes. There are a total of 55 possible attributes, or inputs, determining the single outcome: whether or not the juvenile committed a crime between age 11 and 16. There are 1,131 observations in the data set.

The data used in this project is a random sample taken from the 1997 National Longitudinal Survey of Youth (NLSY97), as administered by the Bureau of Labor Statistics in the U.S. Department of Labor.[1] The NLSY97 is a nationally representative sample of approximately 9,000 young men and women who were 12-16 years old as of December 31, 1996. The youth participants and their parents were both interviewed in the first round of this survey in 1997, with subsequent annual follow-ups for the youths. Note that despite the fact the NLSY97 contains survey responses from parents, only the survey responses from the youth were used for the input variables - the various components of *Nature*, *Nurture*, and *Nerd* as described in the previous paragraph - in these models.

A select few of the variables included in this analysis were appropriately derived from that initial data set, most notably the outcome variable of **crime**. We created the **crime** variable by combining information in six other variables from NLSY97 data that listed the youths' participation in different categories of crime at separate

[1] Accessible at https://www.bls.gov/nls/nlsy97.htm.

ages between 11 and 16. In other words, there were six variables according to age (labeled in NLSY97 as vcrime11, vcrime12, ..., vcrime16) that listed the number of different types of crimes that the youth was involved in. The types were vandalism, petty theft, major theft, motor vehicle theft, property crime, aggravated assault, and selling drugs. Unfortunately, the data was not set up to explain how many crimes within each type were committed. An example should suffice to show the limitation in this data. Assume vcrime13 = 4. That means that the individual committed at least one crime, but perhaps more, within four of the types listed above at age 13. He could have committed one major theft, one petty theft, one vandalism, and one aggravated assault. Or he could have committed 20 major thefts, 10 vandalisms, one property crime, and one aggravated assault. All that matters is that he committed at least one crime of four types. For this reason, it made sense to create **crime** as a single binary variable which simply asks whether the youth committed any crime between age 11 and 16. If any of the vcrime variables has a value above 0, then **crime** gets designated as a 1. Otherwise, if the youth never committed any type of crime during that age range, crime gets set to 0.

There are other issue with the data that should be noted. First, the designation of a real-valued attribute in Table 1 does not necessarily mean any possible value is seen in that range. Most of the real-valued attributes from the table are really just meant as any integers in that range, but, when a sufficient number of unique responses were observed, it did not seem prudent in many cases to write out all of the discrete values. In addition, many of the variables towards the end of the list have oddball answers. Perhaps the respondents did not understand the statements properly, or the data is mislabeled. Either way, it is rather implausible to spend 10.5 hours per typical weekday on homework. It is very doubtful that anyone eats dinner with their family on *7 weekdays per week* unless they have rather strange eating habits. And

it is downright impossible to spend 35 hours in a typical weekday watching TV. We considered deleting some of these obviously misleading observations but ultimately left them all in the analysis. It was difficult to determine whether one value was too much more absurd than its closest neighbor to justify one being kept and the other dropped. For example, 10.5 hours per weekday on homework is quite implausible, but what about 10 hours? Is there a significant difference between 7.5 hours and 7 hours? The cutoffs for these values would have been too arbitrary to justify by quick look inspection. And ultimately, if responses are widely varied and systemically outrageous, presumably the predictive power of those responses will not be very high for determining criminal behavior. In that case, they will not be included in the `Best Predictors` group, which we ultimately view as the most important result of the research.

4 Model Based Learning

We are using the model based learning approach to understand what factors are most important in predicting whether a youth will commit a crime. The output variable **crime** is discrete, taking on the value 1 if the youth committed a crime between ages 11 and 16, and 0 otherwise. We use naive Bayes and Bayesian network classifications in Weka to produce our results. Using our classification of A, U, and E from Figure 1, we check whether *Nature*, *Nurture*, or *Nerd* variables are better predictors of criminal behavior. In each case, we follow the standard approach of using the first 2/3 of the data set to train the model and the remaining 1/3 to test the model. We evaluate the resulting number of correctly and incorrectly classified instances of the test set and the related confusion matrix. In a public policy framework, the variables that have the most power to predict crime should receive the bulk of public money aimed

at preventing criminal behavior. In other words, if *Nurture* variables are much better predictors of crime commission than *Nerd*, perhaps more public money should be spent on parent training courses than on after-school programs.

4.1 Naive Bayes

The naive Bayes classifier makes a simplifying assumption: independence among input variables or attributes. This classifier usually does surprisingly well despite the strength of this assumption. Much evidence has been found suggesting that the naive Bayes classifier is comparable to Bayesian networks and decision trees in performance and accuracy, often showing better speeds (Langley, Iba, and Thomas, 1992; Pazzani, 1996; Patil and Sherekar, 2013).

This method requires the estimation of fewer parameters (we are not estimating n-dimensional variance-covariance matrices between every single attribute). Therefore, the probability of belonging to a class (either committing a crime or not) is calculated from the joint distribution of independent realizations of each attribute. Thus, naive Bayes classifiers assume that the effect of an attribute is independent of the values of other attributes. This assumption is called class conditional independence. The naive Bayes classifier uses Bayes theorem to compute the probability for every observation in each class, given the input attributes. Applying this theorem on a subset of the data allows us to learn a probabilistic model for each class. The calculated posterior probability that a data record belongs to a class C_j is:

$$Pr(C_j|x_i) = \frac{Pr(x_i|C_j)Pr(C_j)}{\sum_{C_k} Pr(x_i|C_k)Pr(C_k)} \tag{1}$$

The prior likelihoods can be obtained from asking an expert or learning them from the dataset. We follow the latter approach. Because of the independence assump-

tion, we can simply multiply the individual conditional probabilities. To estimate these probabilities, we simply calculate shares in the case of discrete attributes. For continuous attributes, we assume a Gaussian distribution, estimate its variance and mean parameters from its sample moments, and use them to calculate the probability or likelihood of observing an attribute take a value in the observed range given the class. The naive Bayes results appear in Section 5.1.

4.2 Bayesian Networks

Bayesian network classifiers do not assume the conditional independence that naive Bayes does. We lose the simplicity of interpretation and calculation that makes naive Bayes so appealing. However, in exchange for the extra complexity, we obtain a model of the relationships between the input variables. Since we are not assuming they are independent, we can show how dependent those attributes are on each other. We get a better idea of how some attributes affect the probability of an observation belonging to a certain class, and exactly through which variables these attributes are affecting that probability. This is desirable especially if we intend to draw policy conclusions from our analyses.

To implement the Bayes network, we calculate joint densities for the set of attributes. We also calculate the conditional dependency between attributes. Any attribute that is *not* a descendant of another in the network is conditionally independent of it (conditional on knowing the class and parents). For each attribute, we are choosing a minimal subset of parents such that the attribute is conditionally independent of any attribute that is not its parent. Whether any two variables are conditionally independent can be deduced from a Bayes network using d-separation.

Additionally, most programs that implement this classifier use these probabilities

to generate a graph that visually displays the correlational and generational structure of the data. This method at least gives us a suggested direction of causality that is not obtained in naive Bayes. In the Bayes network we are modeling, each node is conditionally independent of its non-descendents given its parents. Just as with naive Bayes, the Bayesian network classifier can be used to build Bayesian networks from either expert input or learned probabilities from the data. For our analyses, we choose the latter approach and again use a 2/3-1/3 data split. Given a Bayesian network structure and a training dataset, we can learn the parameters of each node by maximum likelihood. Then, from the different possible solutions, optimization methods of state-space search (like simulated annealing with multiple restarts) finds the Bayesian network structure that best fits the data.

We specify a maximum of 5 parents for the network graphs. These show a compact representation of the joint probability distribution and only require the storage of a number of probabilities exponential in the number of parents per node, not the total number of nodes.

5 Results

5.1 Classifier Performance

Tables 3 and 4 show the performance of the two classifiers. We are most interested in the resulting number of correctly and incorrectly classified instances in each category. The "Guess crime = 1" line shows the success rate if we classified all instances as **crime** = 1. It serves as a baseline comparison, since it is effectively as naive as possible (more so even than the somewhat inappropriately named *naive* Bayes!) as it uses no input variable information but simply assumes **crime** = 1 for every observation since

that is the modal outcome. `Best Predictors` refers to a subset of attributes that were chosen because they were the strongest predictors of the class.[2] The results in both tables include the difference between the correctly/incorrectly classified instances and the "Guess crime = 1" result in square brackets.

The naive Bayes classifier outperformed the Bayes net overall. In both cases, the *Nature* variables were particularly bad predictors of **crime**, performing significantly worse than Guess. Likewise by both classifiers, the *Nerd* variables were the best of the three groups, and even outperformed the predictions that used all variables. Using the `Best Predictors` in a naive Bayes classifier resulted in the greatest predictive power on **crime**, improving over the Guess results by almost 6%.

5.2 Bayesian Network Graphs

The Bayesian network graphs are difficult to present visually when there are many attributes included. We show the Bayesian network graph for the *Nurture* category in Figure 1 as an example. In this graph, we can see that the arrangement of parents-sons in the graph follows what we would expect intuitively (for example, infrequently eating dinner at home increases the probability of committing a crime by a youth directly, and also through the effect it has on how often the youth helps with household chores, showing an effect of actually helping out and spending more time at home).

In Figure 2, we show the Bayesian network graph that determined exactly which variables are included in `Best Predictors` - namely, male, mobile, mattach, mvirtl, nlate, grades8, nfight, schpos, and susp.[3] Using Tables 1 and 2, we see that these nine

[2]As determined by evaluating the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (Hall, 1998).

[3]Male is a binary variable denoting gender. Mobile is a count of the number of residences the youth had since age 12. Mattach is a measure of attachment between the youth and mother. Mvirtl

variables include two *Nature*, two *Nurture*, and five *Nerd* from our original category designations.

5.3 Conditional Probabilities

Finally, a few example tables for the `Best Predictors` are shown in Tables 5-8. These results are one reason why the Bayesian network can be more useful than the naive Bayes analysis, even if the number of parameters that it must estimate can increase the error rate and therefore lead to a lower number of correctly classified instances (as found in our results). In this case, we obtain the exact conditional probabilities of each node in the Bayesian network graphs. These results follow the intuitive direction of causality in most instances. However, the information obtained from these tables goes well beyond intuition. The actual conditional probabilities quantify the extent to which each attribute affects the class crime compared to others.

6 Conclusions

Individually, none of these categories perform well in predicting **crime**, but the *Nerd* attributes performed better at predicting **crime** than *Nurture* and *Nature*. Nearly 60% of the youths remaining in the test set committed a crime. So a completely naive "guessing" approach which always predicted "crime = 1" would be right more often than either classifier using the *Nature* input variables and more often than

is a measure of how much the mother monitored the youth's activities. Nlate is a count of the unexcused late arrivals to school in the previous year. Grades8 is a measure of course grades in 8th grade. Schpos is a measure of how favorably the youth viewed his school. Susp is a binary variable indicating whether the youth had ever been suspended from school. More extensive descriptions can be found in Tables 1 and 2.

the Bayesian network classifier using the *Nurture* input variables. This is somewhat disappointing in that it potentially means that knowing nothing about the individuals is as helpful as knowing entire categories about them. If the cost of obtaining the relevant schooling or family background information is high, there seems little reason to exert the effort or expense in doing so. Although the *Nerd* variables performed the best out of the three categories, there is scant evidence that they are a significantly valuable predictor of criminal behavior on their own.

Aggregating the attributes across all categories and using all of them to predict **crime** also yields disappointing results. *Nerd* actually outperforms the entire input data set using both classifiers, and, as explained in the previous paragraph, the actual predictive power of *Nerd* is not terribly high in the first place. Including all of the attributes only served to further obfuscate whatever little ability the *Nerd* variables did have at explaining criminal behavior in youth.

Overall, naive Bayes outperformed Bayesian networks. We expected Bayesian networks to significantly outperform naive Bayes models in this context since the naive Bayes assumption of independence is so obviously violated, but this was not the case. The Bayesian network predictions were quite a bit worse than the naive Bayes counterparts for *Nurture* and `Best Predictors`, while they were barely better or the same for All Attributes, *Nature* (which performed very poorly in both settings), and *Nerd*. We believe this stems from a lack of sufficient sample size to estimate the additional parameters for Bayesian networks. Too few data records can then lead to poor parameter estimates and subsequent lower classification accuracy. The advantages of Bayesian networks are the conditional probabilities and network structure produced.

Selecting the `Best Predictors` and using them to predict **crime** does improve the model significantly over the categorization approach. The optimally chosen nine attributes come from all three categories, although the categories are not equally

represented. Since *Nerd* did better in the individual category crime predictions, it is not surprising to see the majority of `Best Predictors` attributes come from there. Overall, the `Best Predictors` are able to correctly predict **crime** in 65.5% of the observations. This is a reasonably large improvement over the random guess method at predicting which youths will commit a crime between age 11 and 16.

7 Policy Implications

Given that *Nerd* was the best category of attributes and that within the `Best Attributes` set the *Nerd* category was over-represented, we can conclude that policies aimed towards improving educational variables are paramount at reducing juvenile crime. Policies that ensure that children have opportunities to improve their performance at school will lower their probability of committing a crime. Policies that improve test performance and study habits (such as increasing the number of hours dedicated to doing homework and increasing participation in extra lessons) are recommended.

However, it is also true that one of the main policy lessons that could be drawn from this study is that a single direction policy will be a flawed approach: it is preferable to address aspects in all categories, even if educational policies receive special attention. For example, the `Best Predictor` attributes included variables from each category.

References

Baumgartner, K., Ferrari, S., and Palermo, G. (2008). Constructing Bayesian networks for criminal profiling from limited data. *Knowledge-Based Systems*,

21(7), 563-572.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, A., and York, R. (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.

Deming, D. (2011). Better schools, less crime? *Quarterly Journal of Economics*, 126(4), 2063-2115.

DeAngelis, C. and Wolf, P. (2016). The school choice voucher: A get out of jail card? EDRE Working Paper 01.

Freeman, R. (1999). The economics of crime. *Handbook of Labor Economics*, 3(52), 3529-3571.

Galton, F. (1895). *English Men of Science: Their Nature and Nurture.* D. Appleton.

Ge, X., Conger, R., Cadoret, R., Neiderhiser, J., Yates, W., Troughton, E., and Stewart, M. (1996). The developmental interface between nature and nurture: A mutual influence model of child antisocial behavior and parent behaviors. *Developmental Psychology*, 32(4), 574-589.

Hall, M. (1998). *Correlation-Based Feature Subset Selection for Machine Learning.* Hamilton, New Zealand.

Hall, M. (2007). A decision tree-based attribute weighting filter for naive Bayes. *Knowledge-Based Systems*, 20(2), 120-126.

Herrnstein, R. and Wilson, J. (1985). *Crime and Human Nature: The Definitive Study of the Causes of Crime.* The Free Press. New York.

Hochschild, J. (2003). Social class in public schools. *Journal of Social Issues*, 59(4), 821-840.

Langley, P., Iba, W., and Thomas, K. (1992). An analysis of Bayesian classifiers. In *Proceedings of the Tenth National Conference of Artifical Intelligence*. AAAI Press. 223-228.

Liao, R., Wang, X., Li, L., and Qin, Z. (2010). A novel serial crime prediction model based on Bayesian learning theory. In *Proceedings of International Conference on Machine Learning and Cybernetics*. Qingdao, China.

Patil, T., and Sherekar, S. (2013). Performance analysis of naive Bayes and J48 classification algorithm for data classification. *International Journal of Computer Science and Application*, 6(2), 256-261.

Pazzani, M. (1996). Search for dependencies in Bayesian classifiers. In Fisher, D., and Lenz, H., eds., *Learning from Data: Artificial Intelligence and Statistics*. Springer Verlag.

Raphael, S. and Sills, M. (2006). Urban crime, race, and the criminal justice system in the United States. In Arnott, J., and McMillen, D., eds., *Companion to Urban Economics*. Blackwell Publishing. Oxford, UK.

Reif, A., Reisler, M., Freitag, C., Schneider, M., Eujen, A., Kissling, C., Wenzler, D., Jacob, C., Retz-Junginger, P., Thome, J., Lesch, K., and Retz, W. (2007). Nature and nurture predispose to violent behavior: Serotonergic genes and adverse childhood environment. *Neuropsychopharmacology*, 32, 2375-2383.

Sampson, R. and Laub, J. (1990). Crime and deviance over the life course: The salience of adult social bonds. *American Sociological Review*, 55(5), 609-627.

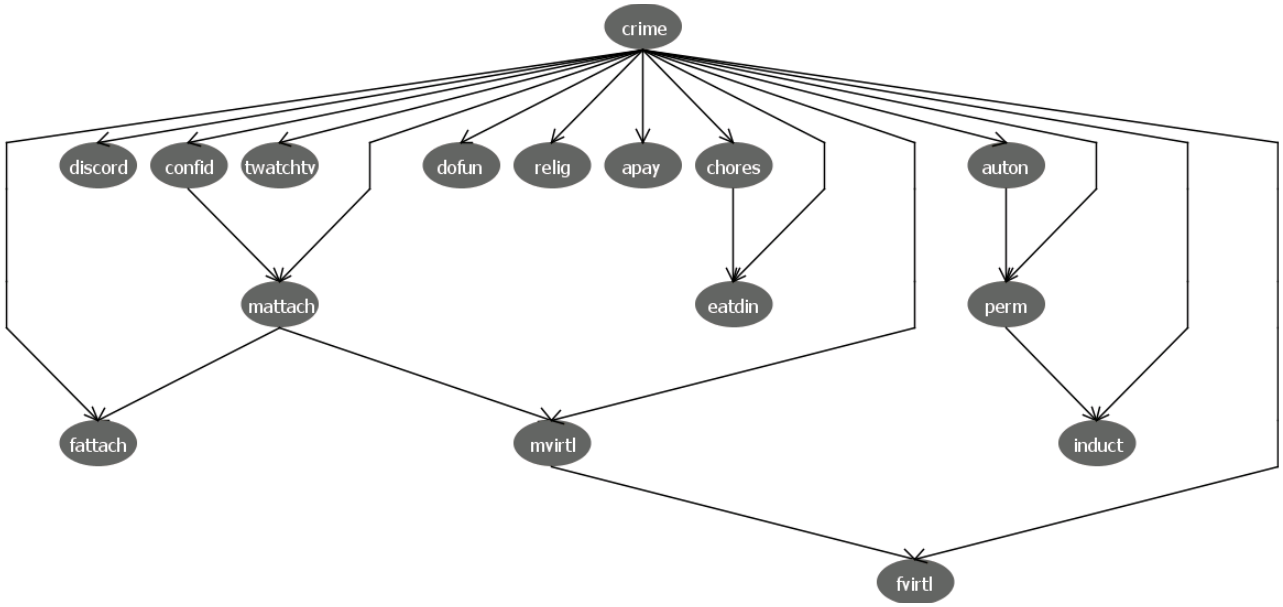

Figure 1: Bayes Network - *Nurture*

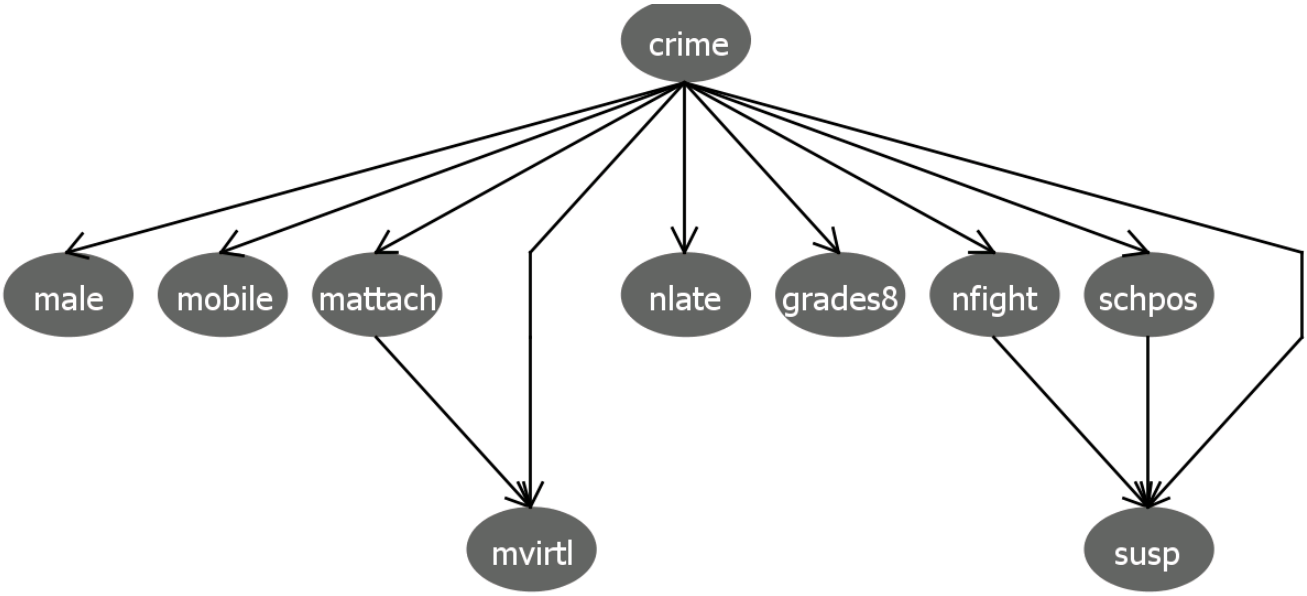

Figure 2: Bayes Network - `Best Predictors`

Table 1: Data Description

Variable Name	Description	Range	Group
MALE	1 if male	Binary	Nature (A)
RACE	1 if white, 2 if black, 3 if Hispanic, 4 if other	{1, 2, 3, 4}	A
WHITE	1 if white	Binary	A
AGE97	Age in 1997	{12, 13}	A
REGION	1 if northeast, 2 if midwest, 3 if south, 4 if west	{1, 2, 3, 4}	A
MOBILE	Average # of different residences each year since age 12	(0.24, 4.67)	A
HHSIZE	Count of # of people in household	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 16}	A
DEADP	1 if at least one of the youth's biological parents is deceased	Binary	A
BBIO	1 if youth resides with both biological parents	Binary	A
STEP	1 if youth resides with one biological parent and step parent	Binary	A
BIOMOM	1 if youth resides with biological mother only	Binary	A
BIODAD	1 if youth resides with biological father only	Binary	A
OTHFAM	1 if youth has some other family living arrangement	Binary	A
MEMP	1 if mother figure employed	Binary	A
FEMP	1 if father figure employed	Binary	A
MLESS	1 if mother figure did not finish high school	Binary	A
MHIGH	1 if mother figure completed high school but has no further education	Binary	A
MCOLL	1 if mother figure has four years of college education	Binary	A
MGRAD	1 if mother figure has over four years of college education	Binary	A
FLESS	1 if father figure did not finish high school	Binary	A
FHIGH	1 if father figure completed high school but has no further education	Binary	A
FCOLL	1 if father figure has four years of college education	Binary	A
FGRAD	1 if father figure has over four years of college education	Binary	A
LOWED	1 if at least one parent figure did not finish high school	Binary	A
PASSETS	Count of # of different assets owned by parents (e.g., property, stocks, car)	{0, 1, 2, 3, 4, 5, 6, 7, 8}	A
INCOME	Household income	(0, 115,000)	A
TEENMOM	1 if biological mother was 19 or younger at first birth	Binary	A
FBORNPAR	1 if at least one parent figure was not born in the US	Binary	A
CONFID	1 if youth turns to parents first w/ emotional or relationship problems	Binary	Nurture (U)
MATTACH	Count of # of statements about mother figure with which youth "agrees" or "strongly agrees" (e.g., I think highly of her, she is a person I want to be like; note that higher scores denote stronger attachment)	{0, 1, 2, 3, 4, 5, 6, 7, 8}	U
MVIRTL	Count of # of people in youth's life about which mother figure knows most things (e.g., who close friends are, who am with when I am not at home; note that higher scores denote stronger monitoring of youth's activities)	{0, 1, 2, 3, 4}	U
FATTACH	Count of # of statements about father figure with which youth "agrees" or "strongly agrees"	{0, 1, 2, 3, 4, 5, 6, 7, 8}	U
FVIRTL	Count of # of people in youth's life about which father figure knows most things	{0, 1, 2, 3, 4}	U
DISCORD	Count of # of statements about parents' usual interaction style (e.g., yell, insult; note that higher scores denote more discord)	{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 11}	U
AUTON	Count of # of behaviors in which youth participates in setting limits either jointly or alone with parents (e.g., how late to stay out, what TV shows to watch)	{0, 1, 2, 3}	U
PERM	Count of # of behavioral limits in which parents respond permissively (e.g., ignoring it, letting youth get away with it)	{0, 1, 2, 3}	U
INDUCT	Count of # of behavioral limits in which parents respond inductively (e.g., discussing it calmly, taking away a priviledge)	{0, 1, 2, 3}	U

Table 2: Data Description (Continued)

Variable Name	Description	Range	Group
TWATCHTV	Count of # of hours per weekday youth typically spends watching TV	(0, 35)	U
CHORES	Count of # of days in a week that youth gets housework done	{0, 1, 2, 3, 4, 5, 6, 7}	U
EATDIN	Count of # of weekdays that youth eats dinner with the family	{0, 1, 2, 3, 4, 5, 6, 7}	U
DOFUN	Count of # of weekdays that youth does something fun with the family	{0, 1, 2, 3, 4, 5, 6, 7}	U
RELIG	Count of # of weekdays that youth does something religious with the family	{0, 1, 2, 3, 4, 5, 6, 7}	U
APAY	1 if youth receives an allowance	Binary	U
PIAT	Math percentile on the Peabody Individual Achievement Test	(0, 100)	Nerd (E)
GRADES8	Grades in 8th grade (1 = mostly below D's, 2 = mostly D's, 3 = C's and D's, 4 = mostly C's, 5 = B's and C's, 6 = mostly B's, 7 = A's and B's, 8 = mostly A's	{1, 2, 3, 4, 5, 6, 7, 8}	E
REPT	1 if youth ever repeated a grade	Binary	E
SKIP	1 if youth ever skipped a grade	Binary	E
SUSP	1 if youth was ever suspended from school	Binary	E
NFIGHT	Count of # of times youth got into fight at school	{0, 1, 2, 3, 4, 5, 6}	E
NLATE	Count of # of times youth was late for school without an excuse	(0, 30)	E
NABSNT	Count of # of times youth was absent from school	(0, 35)	E
SCHPOS	Count of # of statements about his school to which youth "agrees" or "strongly agrees" (e.g., teachers are good, students are fairly graded)	{0, 1, 2, 3, 4, 5}	E
THOMEWK	Count of # of hours per weekday youth typically spends doing homework	(0, 10.5)	E
TEXTRA	Count of # of hours per weekday youth typically spends taking extra classes or lessons	(0, 7)	E
TREAD	Count of # of hours per weekday youth typically spends reading for pleasure	(0, 12)	E

Table 3: Naive Bayes

Category	Correctly Classified Instances	Incorrectly Classified Instances
Guess Crime = 1	229 (59.5) [0]	156 (40.5) [0]
Nature	210 (54.5) [-19]	175 (45.5) [+19]
Nurture	232 (60.3) [+3]	153 (39.7) [-3]
Nerd	235 (61.0) [+6]	150 (39.0) [-6]
All Attributes	233 (60.5) [+4]	152 (39.5) [-4]
`Best Predictors`	252 (65.5) [+23]	133 (34.5) [-23]

Percentages are reported in parentheses ()

Increase of correct predictions over Guess are reported in brackets []

Table 4: Bayes Network

Category	Correctly Classified Instances	Incorrectly Classified Instances
Guess Crime = 1	229 (59.5) [0]	156 (40.5) [0]
Nature	210 (54.5) [-19]	175 (45.5) [+19]
Nurture	220 (57.1) [-9]	165 (42.9) [+9]
Nerd	236 (61.3) [+7]	149 (38.7) [-7]
All Attributes	234 (60.8) [+5]	151 (39.2) [-5]
`Best Predictors`	247 (64.2) [+18]	138 (35.8) [-18]

Percentages are reported in parentheses ()

Increase of correct predictions over Guess are reported in brackets []

Table 5: Conditional Probabilities from Bayesian Network Graph

		Mvirtl				
Crime	**Mattach**	**0**	**1**	**2**	**3**	**4**
0	**0**	0.20	0.20	0.20	0.20	0.20
0	**1**	0.38	0.23	0.23	0.08	0.08
0	**2**	0.26	0.26	0.26	0.19	0.04
0	**3**	0.26	0.26	0.26	0.26	0.11
0	**4**	0.03	0.16	0.16	0.23	0.16
0	**5**	0.06	0.22	0.22	0.24	0.29
0	**6**	0.04	0.10	0.10	0.31	0.22
0	**7**	0.07	0.10	0.10	0.33	0.30
0	**8**	0.02	0.05	0.05	0.26	0.56
1	**1**	0.42	0.23	0.23	0.10	0.03
1	**2**	0.38	0.13	0.13	0.08	0.08
1	**3**	0.29	0.22	0.22	0.16	0.09
1	**4**	0.29	0.35	0.35	0.08	0.08
1	**5**	0.17	0.14	0.14	0.18	0.10
1	**6**	0.09	0.30	0.30	0.18	0.17
1	**7**	0.01	0.12	0.12	0.30	0.29
1	**8**	0.05	0.09	0.09	0.30	0.40

Table 6: Conditional Probabilities from Bayesian Network Graph

	Male	
Crime	**0**	**1**
0	0.63	0.37
1	0.46	0.54

Table 7: Conditional Probabilities from Bayesian Network Graph

	grades 8							
crime	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	0.01	0.01	0.03	0.13	0.17	0.14	0.29	0.22
1	0.02	0.04	0.1	0.13	0.22	0.15	0.24	0.11