Nomograms for Visualization of Naive Bayesian Classifier

Martin Možina[1], Janez Demšar[1], Michael Kattan[2], and Blaž Zupan[1,3]

[1] Faculty of Computer and Information Science, University of Ljubljana, Slovenia
[2] Memorial Sloan Kettering Cancer Center, New York, NY, USA
[3] Dept. Mol. and Human Genetics, Baylor College of Medicine, Houston, TX, USA

{martin.mozina|janez.demsar}@fri.uni-lj.si
kattanm@mskcc.org, blaz.zupan@fri.uni-lj.si

Abstract. Besides good predictive performance, the naive Bayesian classifier can also offer a valuable insight into the structure of the training data and effects of the attributes on the class probabilities. This structure may be effectively revealed through visualization of the classifier. We propose a new way to visualize the naive Bayesian model in the form of a nomogram. The advantages of the proposed method are simplicity of presentation, clear display of the effects of individual attribute values, and visualization of confidence intervals. Nomograms are intuitive and when used for decision support can provide a visual explanation of predicted probabilities. And finally, with a nomogram, a naive Bayesian model can be printed out and used for probability prediction without the use of computer or calculator.

1 Introduction

Compared to other supervised machine learning methods, naive Bayesian classifier (NBC) is perhaps one of the simplest yet surprisingly powerful technique to construct predictive models from labelled training sets. Its predictive properties have often been a subject of theoretical and practical studies (*e.g.* [1, 2]), and it has been shown that despite NBC's assumption of conditional independence of attributes given the class, the resulting models are often robust to a degree where they match or even outperform other more complex machine learning methods.

Besides good predictive accuracy, NBC can also provide a valuable insight to the training data by exposing the relations between attribute values and classes. The easiest and the most effective way to present these relations is through visualization. But while the predictive aspect of NBC has been much studied, only a few reports deal with visualization and explanation capabilities of NBC. In this, a notable exception is the work of Kononenko [2] and Becker *et al.* [3]. Kononenko introduced the concept of information that is gained by knowing the value of a particular attribute. When using the NBC for classification, Kononenko's information gains can offer an explanation on how the values of the attributes influenced the predicted probability of the class. Becker and coauthors proposed an alternative approach to visualization of NBC that is also available

as Evidence Visualizer in the commercial data mining suite MineSet. Evidence Visualizer uses pie and bar charts to represent conditional probabilities, and besides visualization of the model offers interactive support to prediction of class probabilities.

In the paper, we propose an alternative method to visualization of a NBC that clearly exposes the quantitative information on the effect of attribute values to class probabilities and uses simple graphical objects (points, rulers and lines) that are easier to visualize and comprehend. The method can be used both to reveal the structure of the NBC model, and to support the prediction.

The particular visualization technique we rely on are nomograms. In general, a *nomogram* is any graphical representation of a numerical relationships. Invented by French mathematician Maurice d'Ocagne in 1891, the primary means of a nomogram was to enable the user to graphically compute the outcome of an equation without doing any calculus. Much later, Lubsen and coauthors [4] extended the nomograms to visualize a logistic regression model. They show the utility of such a device on a case for prediction of probability of diagnosis of acute myocardial infarction. Their nomogram was designed so that it can be printed on the paper and easily used by physicians to obtain the probability of diagnosis without resorting to a calculator or a computer. With an excellent implementation of logistic regression nomograms in a Design and hmisc modules for S-Plus and R statistical packages by Harrell [5], the idea has recently been picked up; especially in the field of oncology, there are now a number of nomograms used in daily clinical practice for prognosis of outcomes of different treatments that have been published for variety of cancer types (*e.g.* [6]; see also http://www.baylorcme.org/nomogram/modules.cfm).

Our NBC nomograms use a similar visualization approach to that of Harrell for logistic regression, and well conform to the NBC visualization design principles as stated by Becker [3]. In the paper, we first show how to adapt the NBC to be suitable for visualization with a nomogram. We also propose the means to compute confidence intervals for the contributions of attribute values and for the class probability, and include these in visualization. We discuss on the differences between our visualization approach and that of Evidence Visualizer, and compare NBC nomograms to those for logistic regression. We show that a simple adjustment of an NBC nomogram can support a visual comparison between NBC and logistic regression models. The particular benefits of the approach and ideas for the further work are summarized in the conclusion.

2   Naive Bayesian Nomogram

Let us start with an example. Fig. 1 shows a nomogram for a NBC that models the probability for a passenger to survive the disaster of the HMS Titanic. The nomogram, built from the well-known Titanic data (http://hesweb1.med.virginia. edu/biostat/s/data/), includes three attributes that report on the travelling class (first, second, and third class, or a crew member), age (adult or child), and gender of the passenger.
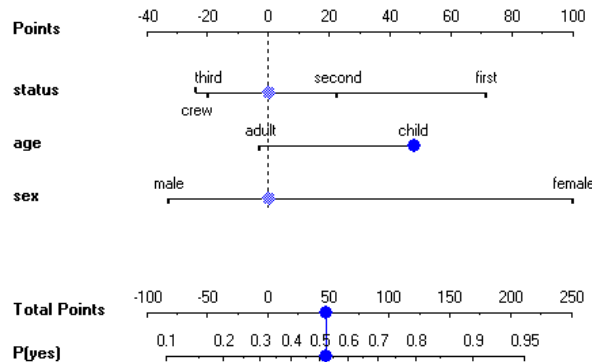
Fig. 1. A nomogram for prediction of survival probability of a passenger on HMS Titanic.

Of 2201 passengers on Titanic, 711 (32.3%) survived. To make a prediction, the contribution of each attribute is measured as a point score (topmost axis in the nomogram), and the individual point scores are summed to determine the probability of survival (bottom two axes of the nomogram). When the value of the attribute is unknown, its contribution is 0 points. Therefore, not knowing anything about the passenger, the total point score is 0, and the corresponding probability equals to the unconditional prior. The nomogram in Fig. 1 shows the case when we know that the passenger is a child; this score is slightly less than 50 points, and increases the posterior probability to about 52%. If we further know that the child travelled in the first class (about 70 points), the points would sum to about 120, with a corresponding probability of survival of about 80%.

Besides enabling the prediction, the naive Bayesian nomogram nicely reveals the structure of the model and the relative influences of the attribute values to the class probability. For the Titanic data set, gender is an attribute with the biggest potential influence on the probability of passenger's survival: being female increases the chances of survival the most (100 points), while being male decreases it (about −30 points). The corresponding line in the nomogram for this attribute is the longest. Of the three attributes age is apparently the least influential, where being a child increases the probability of survival. Most lucky were also the passengers of the first class for which – considering the status only – the probability of survival was much higher than the prior.

In the following we show how to represent the NBC in the way to be applicable for visualization with a nomogram. We further introduce confidence intervals, and discuss several other details on our particular implementation.

2.1 Derivation of Naive Bayesian Nomogram

Naive Bayesian rule to assess the probability of class c given an instance X with a set of attribute values $X = \langle a_i, a_2 \ldots a_n \rangle$ is:

$$P(c|X) = \frac{P(a_i, a_2 \ldots a_n|c)P(c)}{P(X)} = \frac{P(c)\prod_i P(a_i|c)}{P(X)} \tag{1}$$

We call class c a target class, since it will be the one represented in the nomogram. The probability of the alternative class (or alternative classes) $\bar{c}$ is $P(\bar{c}|X)$, and dividing the two we obtain:

$$\text{Odds} = \frac{P(c|X)}{P(\bar{c}|X)} = \frac{P(c)\prod_i P(a_i|c)}{P(\bar{c})\prod_i P(a_i|\bar{c})} \tag{2}$$

In terms of the log odds (logit $P = \log\frac{P}{1-P}$), this equation translates to:

$$\text{logit } P(c|X) = \text{logit } P(c) + \sum_i \log\frac{P(a_i|c)}{P(a_i|\bar{c})} \tag{3}$$

The terms in summation can be expressed as odds ratios (OR):

$$\frac{P(a_i|c)}{P(a_i|\bar{c})} = \frac{\frac{P(c|a_i)}{P(\bar{c}|a_i)}}{\frac{P(c)}{P(\bar{c})}} = \text{OR}(a_i) \tag{4}$$

and estimate the ratio of posterior to prior probability given the attribute value a_i.[4] We now take the right term in (3) and call it $F(c|X)$:

$$F(c|X) = \sum_i \log\frac{P(a_i|c)}{P(a_i|\bar{c})} = \sum_i \log \text{OR}(a_i) \tag{5}$$

and use it for the construction of the central part of the nomogram relating attribute values to point scores. The individual contribution (point score) of each known attribute value in the nomogram is equal to $\log \text{OR}(a_i)$, and to what we have referred as the sum of point scores corresponds to $F(c|X)$.

Using Eq. 5, we can now derive the central part of the nomogram from Fig. 1. As a target class, we will use the survival of the passengers. For each attribute value, we compute the individual contributions (*i.e.* point scores) from the instance counts in Table 1. For example, the log odds ratio for the passenger in the first class is 1.25, as the odds for surviving in the first class are $203/122 = 1.67$, unconditional odds for surviving are $711/1490 = 0.48$, and their log ratio is $\log(1.67/0.48) = 1.25$. Similarly, the log odds ratio for the second-class passengers is $\log\frac{118/167}{0.48} = 0.393$, and for the female passengers is $\log\frac{344/126}{0.48} = 1.744$. Notice that instead of relative frequencies used here, probabilities could also be estimated by other methods, like Laplace or m-estimate [8].

---

[4] Our use of odds ratios is a bit different from that in logistic regression. Instead of relating posterior and prior probability, odd ratios in logistic regression relate the odds at the two different values of a binary attribute [7].

Table 1. Number of instances in Titanic data set with a particular value of an attribute and class.

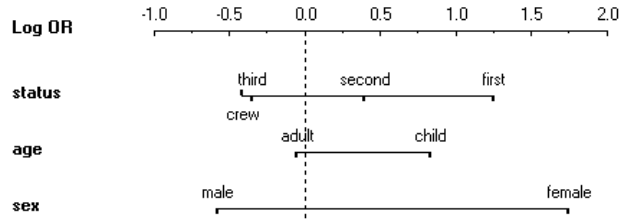| ATTRIBUTE	VALUE	CLASS=YES	CLASS=NO
STATUS	FIRST	203	122
	SECOND	118	167
	THIRD	178	528
	CREW	212	673
AGE	ADULT	654	1438
	CHILD	57	52
SEX	MALE	367	1364
	FEMALE	344	126

Fig. 2. A central part of the Titanic nomogram showing the log odds ratios for survival given different values of the attributes.

These and all other log odds ratios for different attribute values form the central part of the nomogram shown in Fig. 2. This is equal to the corresponding part of the nomogram from Fig. 1, except that for the latter we have re-scaled the units so that log odds ratio of 1.744 – a maximal absolute log odds ratio in the nomogram – represents 100 points. Log odds ratio is a concept that experts, like those from biomedical statistics, do understand and can interpret. For others, a scale with points from -100 to 100 may provide more comfort, and summing-up the integers may be easier than using real numbers. Also, it may be easier to compare the contributions of different attribute values in 100 points scale.

For the part of the nomogram that relates the sum of individual point scores to the class probability, we start from Eqs. 3 and 5 and obtain:

$$\log \frac{P(c|X)}{1 - P(c|X)} = \log \frac{P(c)}{1 - P(c)} + F(c|X) \tag{6}$$

From this, we compute the probability $P(c|X)$ as:

$$P(c|X) = [1 + e^{-\log P(c)/(1-P(c)) - F(c|X)}]^{-1} \tag{7}$$

The lower part of the nomogram, which relates the sum of points as contributed by the known attributes to the class probability, is then a tabulation of a function $P(c|X) = f[F(c|X)]$. For our Titanic example, this part of the nomogram is shown in Fig. 3.
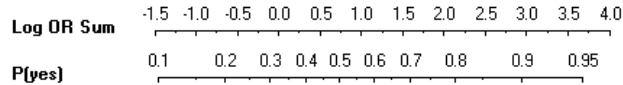
Fig. 3. The part of Titanic nomogram to determine the the probability of survival from the sum of log odds ratios as contributed from the known attribute values.

2.2 Confidence Intervals

The point scores in the nomogram, *e.g.* the odds ratios $\text{OR}(a_i)$, are estimated from the training data. It may be therefore important for the user of the nomogram to know how much to trust these estimates. We can provide this information through confidence intervals.

The $1 - \alpha$ confidence intervals of $\widehat{\text{OR}}(a_i)$ are estimated as (see [7])

$$\widehat{\text{OR}}(a_i) \pm z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}(\widehat{\text{OR}}(a_i))} \tag{8}$$

where $\widehat{\text{Var}}(\cdot)$ is computed as: [5]

$$\widehat{\text{Var}}(\text{logit } \hat{P}(c)) = [N\hat{P}(c)\hat{P}(\bar{c})]^{-1} \tag{9}$$

$$\widehat{\text{Var}}(\widehat{\text{OR}}(a_i)) = [N_{a_i}\hat{P}(c|a_i)\hat{P}(\bar{c}|a_i)]^{-1} - \widehat{\text{Var}}(\text{logit } \hat{P}(c)) \tag{10}$$

where N is a number of training examples, and N_{a_i} is a number of training examples that include attribute value a_i.

Fig. 4 shows a Titanic survival nomogram that includes the confidence intervals for $\alpha = 0.95$.

2.3 Implementation

We have implemented a NBC nomogram as a widget within a machine learning suite Orange [9]. The widget (see Fig. 4 for a snapshot) supports visualization of a nomogram, use of confidence intervals, and can, for each attribute value, plot a bar with height proportional to the number of particular instances.

Our implementation supports the classification. Attribute values (dots on attribute axis) can be moved across the axis, where we can also select values between two value marks (weighted distributions). Class probabilities and associated confidence intervals are updated instantaneously with any change in the data or corresponding naive Bayesian model.

---

[5] Although we regard the computation of the confidence intervals for NBC as an important new addition for this method, our paper focuses on visualization of the confidence intervals and we omit the proof and derivation of confidence intervals due to space considerations.
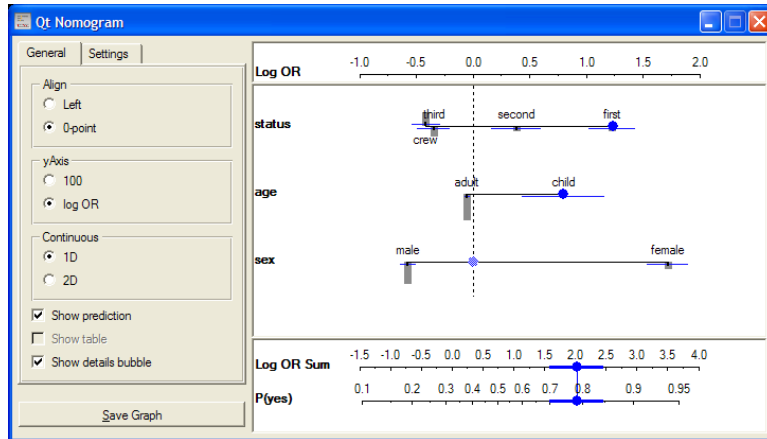
Fig. 4. Orange widget with the Titanic nomogram that includes confidence intervals for contributions of attribute values and class probabilities. For a woman travelling in the first class, the probability of survival is with 95% confidence between 0.87 and 0.92.

3 Discussion and Related Work

Naive Bayesian nomograms, as presented in this paper, are a visualization technique that we adapted and extended from logistic regression [4, 5]. In this, we have considered the design requirements for visualization of NBC as proposed by Becker *et al.* Below, we briefly review them and point out how we have addressed them. In the review of related work, we make a note on the explanation technique for NBC as proposed by Kononenko [2]. We then discuss the differences of our method with that of Evidence Visualizer [3]. Finally, we consider the related work on logistic regression nomograms, discuss the differences and highlight the advantages of naive Bayesian nomograms.

3.1 Design Principles

Becker *et al.* [3] list a number of design requirements for a visualization of a naive Bayesian classifier. With naive Bayesian nomograms, we address all of them. In particular, because of their simplicity, visualized NBC models should be easily understood by non-experts. In the nomogram, comparing the span of an attribute axis easily identifies the important attributes. The effects of each attribute value are also clearly represented in a nomogram, making it easy to spot a direction and magnitude of the influence. Attribute axis are aligned to zero-point influence (prior probability), which allows for a straightforward comparison of contributions across different values and attributes. Confidence intervals, in addition to histograms with record counts, inform the user about reliability of the estimates presented in the nomograms. Nomogram may also be a practical visualization method when dealing with larger number of attributes. For example, sorting attributes by their impact (*e.g.* highest overall/positive/negative influence) offers a means to study and gain a valuable insight into a large naive Bayesian models.

Nomogram-based class-characterization is particularly straightforward for domains with binary class: the zero-point influence line vertically splits the nomogram to the right (positive) and left (negative) part. The visualized class is characterized with the attribute values on the right, whereas the other class is characterized with values presented on the left side of the nomogram. Accordingly, the values farthest from the center are the most influential class indicators.

In our implementation, nomograms can be used for interactive, click-and-drag classification and what-if analysis. Alternatively, they can be printed out and used for probability prediction without the use of computer or calculator.

3.2 Related Work on Nomograms and Visualization of Naive Bayesian Classifier

In the early nineties, Kononenko [2] realized that besides good predictive performance NBCs can be used to *explain* to what degree and how attribute values influence the class probability when classifying an example. He showed that NBCs can be written as a sum of information contributed by each attribute, where a contribution of an attribute value a_i is $\log_2 P(c|a_i) - \log_2 P(c)$.

Zupan *et al* [10] plotted these contributions in a form similar to nomograms proposed in this paper. The main disadvantage of their approach, however, is that the partial attribute scores (influences) have to be summed for each class separately, and then normalized.

Despite the popularity of NBCs within machine learning community, there have not been many reports on methods to visualize it, and even fewer are practical implementations of these. A notable exception is Evidence Visualizer implemented within a data mining suite MineSet [3]. Evidence Visualizer offers two different views of NBCs, and we present an example using the Titanic NBC model in Fig. 5. In the pie chart display, there is a pie for each attribute-value combination, with slices corresponding to distinct classes and where the size of each slice is proportional to $P(a_i|c_j)$. The height of the pie is proportional to number of instances with an attribute value a_i. In the bar-chart representation, each bar signifies the evidence for a particular class c_j as proportional to $-\log(1 - P(a_i|c_j))$. Probability confidence intervals are displayed through color saturation. Evidence Visualizer can be used either in an exploratory way or for making predictions.

While models in Evidence Visualizer may be visually attractive, we believe that NBC nomograms as presented in this paper have several advantages. Both methods aim at visualizing the effects the attribute values have on class probability, where Evidence Visualizer uses bar/pie heights, size of pie slices and color saturations while NBC nomogram uses a simpler position-based representation. Visualization theory [11] gives a clear advantage to positional, line-based, visualization of quantitative information as opposed to that using more complex 2-D and 3-D objects. When comparing relative evidences contributed by different attribute values, a positional visualization should be clearer to that of comparing the size of the pie slices or heights of the bars. In particular, pie-charts have been

Fig. 5. Evidence Visualizer, a visualization of NBC from Titanic data set in MineSet.

criticized by Tufte [11] for their failure to order numbers along a visual dimension, and for reasons of poor interpretability of multiple pie charts. Ware [12] also reports that comprehension of absolute quantities when visualized through color saturations may be at least hard, while interpreting the visualization of confidence intervals through lines of variable length in the nomogram should be more straightforward.

A distinct advantage of a nomogram is that it uses simpler graphical objects and can thus be used in, for instance, systems for decision support on smaller-resolution devices such as handhelds or other mobile devices. The shortcoming of nomograms as compared to the Evidence Visualizer is in visualization in case of multiple-classes: the nomograms visualize one class at a time.

Our NBC nomograms stem from the work of Lubsen [4] and Harrell [5] on visualization of logistic regression. Logistic regression models the probability of target class value as:

$$P(c|X) = \frac{1}{1 + e^{-\beta_0 - \sum_i \beta_i x_i}} \tag{11}$$

where x_i is the value of i-th attribute being 0 or 1 in case of a binary attribute. Notice that m-valued nominal attributes can be encoded with $m - 1$ binary dummy variables. Log odds for the above probability is:

$$\text{logit } P(c|X) = \beta_0 + \sum_i \beta_i x_i \tag{12}$$

The position of a particular attribute value x_i in the logistic regression nomogram is determined through the product $\beta_i x_i$. In such nomogram, one of the values of the attributes will always be displayed at point 0, whereas others will span to the right or left depending on the sign of β. Adhering to the particular (and well-known) implementation by Harrell [5], all attributes are presented so that the leftmost value is drawn at the point 0, and the effects of aligning the attribute axis in this way are compensated through the appropriate change in β_0. The lower part of the nomogram (determination of class probability from the sum of attribute value points) is derived in a similar way as for NBC nomograms (see Fig. 6.a for a nomogram of logistic regression model for the Titanic data set).

There are several important differences between logistic and NBC nomograms, all stemming from the differences between the two modelling methods.
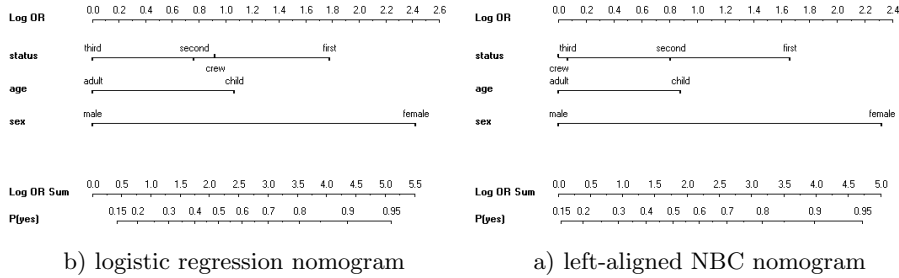
Fig. 6. Comparison of the nomograms for the Titanic data set.

NBC nomograms depict both negative and positive influences of the values of attributes, and differently from logistic regression nomograms, do not anchor one of the values to the 0-point. In this sense, logistic regression nomograms are less appropriate for class characterization. To illustrate this point, values that appear on logistic regression nomogram at 0-point are shown as having the same effect on the class probability, while they may be found at completely different positions in corresponding NBC nomograms. In that, we believe an NBC nomogram offers a better insight into the modelled domain. Another advantage is handling of unknown values. Namely, one needs to specify all attributes when reasoning with a logistic regression nomogram, while NBC nomograms offer a nice interactive one-value-at-a-time update of class probabilities.

To compare NBC and logistic regression nomograms, we can alter the presentation of the NBC nomogram so that the leftmost values for each of the attribute are aligned and their log odds ratio set to 0. We call this a *left-aligned NBC nomogram*. Alignment changes, *e.g.* offsets of the axis for each of the attributes, are reflected in an appropriate update of the lower part of the nomogram.

Fig. 6.a shows a left-aligned nomogram for the Titanic data set, and compares it to the logistic regression nomogram. The two nomograms look very alike, and the effect of the attributes is comparable even on the absolute, log odds ratio scale. The only noticeable difference is a position of the crew value of the attribute status. To analyze this, we used interaction analysis [13] which showed that this attribute strongly interacts with age (crew members are adults) and sex (most of them are male). It seems that the conditional attribute independence assumption of the NBC is most violated for this attribute and value, and hence the difference with logistic regression which is known to be able to compensate for the effects of attribute dependencies [7].

Fig. 7 shows a comparison of the two nomograms for another data set called voting. This is a data set from a UCI Machine Learning Repository [14] on sixteen key votes of each of the U.S. House of Representatives Congressmen of which we have selected six most informative votes for the figure). Visual comparison of the two nomograms reveals some similarities (the first three attributes) and quite some differences (the last three). As for Titanic, we have also noticed that attribute interaction analysis can help explain the differences between the two nomograms.
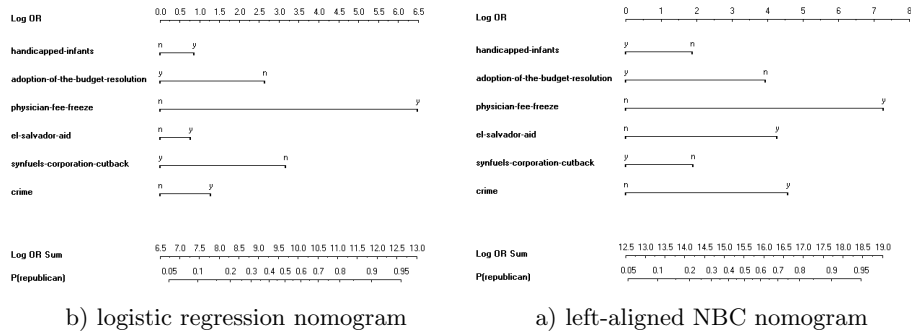
b) logistic regression nomogram   a) left-aligned NBC nomogram

Fig. 7. Comparison of the nomograms for the voting data set.

It is beyond this paper to compare NBC and logistic regression, which has otherwise received quite some attention recently [15]. With the above examples, however, we wanted to point out that nomograms may be the right tool for experimental comparison of different models and modelling techniques, as it allows to easily spot the similarities and differences in the structure of the model.

4 Conclusion

In words of Colin Ware, "one of the greatest benefits of data visualization is the sheer quantity of information that can be rapidly interpreted if it is presented well" [12]. As the naive Bayesian classifier can be regarded as a simple yet powerful model to summarize the effects of attributes for some classification problems, it is also important to be able to clearly present the model to comprehend these effects and gain insight to the data.

In this paper, we show how we can adapt naive Bayesian classifiers and present them with a well established visualization technique called nomograms [5, 6]. The main benefit of this approach is simple and clear visualization of the complete model and the quantitative information it contains. The visualization can be used for exploratory analysis and decision making (classification), and we also show that it can be used effectively to compare different models, including those coming from logistic regression.

There are several aspects of naive Bayesian nomograms that deserve further attention and investigation. In the paper, our examples are all binary classification problems. Nomograms are intended to visualize the probability of one class against all others, and in principle for non-binary classification problems one would need to analyze several nomograms. Also, we have limited the scope of this paper to the analysis and presentation of data sets that include only nominal attributes. In principle, and as with logistic regression nomograms [5], one can easily present continuous relations, and we are now extending naive Bayesian nomograms in this way.

Acknowledgement

References

1. Domingos, P., Pazzani, M.: Beyond independence: conditions for the optimality of the simple Bayesian classifier. In: Proceedings of the Thirteenth International Conference on Machine Learning, Bari, Italy, Morgan Kaufmann (1996) 105–112
2. Kononenko, I.: Inductive and bayesian learning in medical diagnosis. Applied Artificial Intelligence **7** (1993) 317–337
3. Becker, B., Kohavi, R., Sommerfield, D.: Visualizing the simple Bayesian classifier. In Fayyad, U., Grinstein, G., Wierse, A., eds.: Information Visualization in Data Mining and Knowledge Discovery. Morgan Kaufmann Publishers, San Francisco (2001) 237–249
4. Lubsen, J., Pool, J., van der Does, E.: A practical device for the application of a diagnostic or prognostic function. Methods of Information in Medicine **17** (1978) 127–129
5. Harrell, F.E.: Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer, New York (2001)
6. Kattan, M.W., Eastham, J.A., Stapleton, A.M., Wheeler, T.M., Scardino, P.T.: A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J Natl Cancer Inst **90** (1998) 766–71
7. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression. John Wiley & Sons, New York (2000)
8. Cestnik, B.: Estimating probabilities: A crucial task in machine learning. In: Proceedings of the Ninth European Conference on Artificial Intelligence. (1990) 147–149
9. Demšar, J., Zupan, B.: Orange: From experimental machine learning to interactive data mining. White Paper [`http://www.ailab.si/orange`], Faculty of Computer and Information Science, University of Ljubljana (2004)
10. Zupan, B., Demšar, J., Kattan, M.W., Beck, J.R., Bratko, I.: Machine learning for survival analysis: A case study on recurrence of prostate cancer. Artificial Intelligence in Medicine **20** (2000) 59–75
11. Tufte, E.R.: The visual display of quantitative information. Graphics Press, Cheshire, Connecticut (1983)
12. Ware, C.: Information Visualization: Perception for Design. Morgan Kaufmann Publishers (2000)
13. Jakulin, A., Bratko, I.: Analyzing attribute dependencies. In: Proc. of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases, Dubrovnik (2003) 229–240
14. Murphy, P.M., Aha, D.W.: UCI Repository of machine learning databases [`http://www.ics.uci.edu/~mlearn/mlrepository.html`]. Irvine, CA: University of California, Department of Information and Computer Science (1994)
15. Ng, A., Jordan, M.: On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In: Proc. of Neural Information Processing Systems). Volume 15. (2003)