Measuring overlap in logistic regression

Andreas Christmann[1] and Peter J. Rousseeuw[2]

May 03, 1999

[1] University of Dortmund, SFB-475, HRZ, D-44221 Dortmund, Germany
[2] Universitaire Instelling Antwerpen (UIA), Department of Mathematics and Computer Science, Universiteitsplein 1, B-2610 Wilrijk, Belgium

SUMMARY

In this paper we show that the recent notion of regression depth can be used as a data-analytic tool to measure the amount of separation between successes and failures in the binary response framework. Extending this algorithm allows us to compute the overlap in data sets which are commonly fitted by logistic regression models. The overlap is the number of observations that would need to be removed to obtain complete or quasicomplete separation, i.e. the situation where the logistic regression parameters are no longer identifiable and the maximum likelihood estimate does not exist. It turns out that the overlap is often quite small.

Key words: Linear discriminant analysis; Logistic regression; Outliers; Overlap; Probit regression; Regression depth; Separation.

1 Introduction

Logistic regression is used to model the probability that an event occurs, depending on a vector of explanatory variables, say $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p-1}) \in \mathrm{I\!R}^{p-1}$. Often these events can be interpreted as success and failure. The logistic model with an intercept term assumes that the responses y_i are realisations of independent random variables Y_i which are Bernoulli distributed with success probabilities

$$\Lambda((\mathbf{x}_i, 1)\theta') \in (0,1)\,, \quad i = 1, \ldots, n. \tag{1}$$

Here $\Lambda(t) = 1/[1 + \exp(-t)]$ denotes the cumulative distribution function of the logistic distribution, and $\theta \in \mathrm{I\!R}^p$ is unknown. Data sets analyzed with such models have the form $Z_n = \{(x_{i,1}, \ldots, x_{i,p-1}, y_i);\ i = 1, \ldots, n\} \subset \mathrm{I\!R}^p$ where $y_i \in \{0,1\}$ for $i = 1, \ldots, n$. We will always assume that the design matrix has full column rank.

The classical estimator of the unknown parameter vector is the maximum likelihood estimator, c.f. Cox and Snell (1989). However, the maximum likelihood estimate of θ does not always exist. Conditions for its existence were investigated by Albert and Anderson (1984) and Santner and Duffy (1986). They say that the data set is *completely separated* if there exists a vector $\theta \in \mathrm{I\!R}^p$ such that

$$(\mathbf{x}_i, 1)\theta' > 0 \quad \text{if} \quad y_i = 1 \tag{2}$$

$$(\mathbf{x}_i, 1)\theta' < 0 \quad \text{if} \quad y_i = 0 \tag{3}$$

for $i = 1, \ldots, n$. A data set which is not completely separated is *quasicompletely separated* if there exists a vector $\theta \in \mathrm{I\!R}^p \backslash \{0\}$ such that

$$(\mathbf{x}_i, 1)\theta' \geq 0 \quad \text{if} \quad y_i = 1 \tag{4}$$

$$(\mathbf{x}_i, 1)\theta' \leq 0 \quad \text{if} \quad y_i = 0 \tag{5}$$

for all i and if there exists $j \in \{1, ..., n\}$ such that $(\mathbf{x}_j, 1)\theta' = 0$. A data set is said to have *overlap* if there is no complete separation and no quasicomplete separation. Albert and Anderson (1984) and Santner and Duffy (1986) show that the maximum likelihood estimate of θ exists if and only if the data set has overlap. A geometrical interpretation of their result is that the maximum likelihood estimate exists if and only if there is *no* hyperplane which separates successes and failures, where the hyperplane itself may contain both successes and failures.

From a robustness point of view, this yields a problem. Many robust estimators are constructed such that outlying points are deleted or appropriately downweighted. However, it can happen that the whole data set has overlap but the reduced data set does not. In such a situation the robust estimator applied to the whole data set does not exist, see Künsch, Stefanski and Carroll (1989). The latter authors discuss the existence problem and note that it arises regardless of the regression estimator being used, since it is linked to the parametrization of the logistic regression model. In other words, when the data have no overlap the parameters in the logistic model are not identifiable. Künsch, Stefanski and Carroll (1989, p. 466) propose to use their M-estimators with a series of different tuning constants to study the impact of outliers on the estimated parameter vector. The authors concluded that "... it should be checked how close the data are to indeterminacy" and that "... it would be interesting to have other criteria".

The aim of the present paper is to give an answer to these questions by measuring the overlap. We denote by n_{overlap} the smallest number of observations whose removal destroys the overlap of successes and failures. In a logistic regression model, the overlap n_{overlap} is the smallest number of observations that need to be removed to make the maximum likelihood estimate nonexistent. In the same vein, denote by n_{complete} the smallest number of observations whose removal yields complete separation. In other words, this is the minimal number of misclassifications in the training data for *any* linear discriminant function. By definition, always $n_{\text{overlap}} \leq n_{\text{complete}}$.

This paper gives a procedure to determine n_{overlap}, n_{complete}, and the corresponding set(s) of indices corresponding to these observations. Recently, Rousseeuw and Hubert (1999) proposed the regression depth for linear regression models. It will be shown that the regression depth can be used to measure the amount of separation between successes and failures in data sets which are commonly fitted by logistic regression models. Connections between the regression depth and n_{overlap} will also be investigated.

2 Regression depth

In a linear regression model the data set is of the form $Z_n = \{(x_{i,1}, \ldots, x_{i,p-1}, y_i);\ i = 1, \ldots, n\} \subset \mathbb{R}^p$. Denote the $\mathbf{x}-$part of each data point by $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p-1}) \in \mathbb{R}^{p-1}$. The aim is to fit y_i by an affine hyperplane in $\mathbb{R}^p$, i.e. by

$$g((\mathbf{x}_i, 1)\theta') = \theta_1 x_{i,1} + \ldots + \theta_{p-1} x_{i,p-1} + \theta_p \tag{6}$$

where $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$. In this setting, Rousseeuw and Hubert (1999) gave the following two definitions.

Definition 2.1 *A vector $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$ is called a* **nonfit** *to Z_n iff there exists an affine hyperplane V in $\mathbf{x}-$space such that no $\mathbf{x}_i$ belongs to V, and such that the residual $r_i(\theta) = y_i -\ g((\mathbf{x}_i, 1)\theta') > 0$ for all $\mathbf{x}_i$ in one of its open halfspaces, and $r_i(\theta) < 0$ for all $\mathbf{x}_i$ in the other open halfspace.*

Definition 2.2 *The* **regression depth** *of a fit $\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$ relative to a data set $Z_n \subset \mathbb{R}^p$ is the smallest number of observations that need to be removed to make θ a nonfit in the sense of Definition 2.1. Equivalently, rdepth(θ, Z_n) is the smallest number of residuals that need to change sign.*

The regression depth of a fit is invariant with respect to monotone transformations, in the sense that one can replace y_i by $h(y_i)$ where h is a strictly monotone function if the link function g is replaced by $h \circ g$ at the same time. (This is true because the regression depth only depends on the explanatory variables $\mathbf{x}_i$ and the sign of the residuals $r_i(\theta)$.) This invariance property does not hold for the objective function of most regression estimators, such as least squares, least absolute values, and least trimmed squares (Rousseeuw, 1984).

Let us now consider the case of logistic regression for binary response variables. The regression depth can be defined for data sets usually analyzed via logistic regression in the same way as given above, if the cumulative distribution function Λ of the logistic distribution is used instead of g.

From Definition 2.2 it follows for logistic models that the regression depth of a fit θ relative to Z_n is equal to the regression depth of $-\theta$ relative to the data set $\{(x_{i,1}, \ldots, x_{i,p-1}, 1 - y_i);\ i = 1, \ldots, n\}$. Hence, the regression depth is invariant with respect to different codings of the binary response variable.

Let us illustrate the definition of the regression depth by an artificial data set with two explanatory variables x_1 and x_2 and an intercept term:

$$\mathbf{X} = \begin{pmatrix} -1.5, & -1, & 0, & 0, & 1, & 1, & 2, & 3, & 3, & 3.5 \\ 0, & 3, & 1, & 2, & 2, & 4, & 2, & 1, & 3, & 4 \end{pmatrix}', \tag{7}$$

$$\mathbf{y} = (0,\ *,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1)'. \tag{8}$$

If the data point y_2 denoted by $*$ in (8) is a failure, i.e. $y_2 = 0$, then the sets $\{y_i = 0; i = 1, \ldots, n\}$ and $\{y_i = 1; i = 1, \ldots, n\}$ can be separated by an appropriate hyperplane, which is indicated as a line in Figure 1, and hence $n_{\text{overlap}} = n_{\text{complete}} = 0$. The maximum likelihood estimate of θ does not exist in that case, due to complete separation.

If the data point denoted by $*$ in (8) has $y_2 = 1$, then the sets $\{y_i = 0; i = 1, \ldots, n\}$ and $\{y_i = 1; i = 1, \ldots, n\}$ cannot be separated by a hyperplane, and $n_{\text{overlap}} = n_{\text{complete}} = 1$. In that case, the maximum likelihood estimate of θ does exist.

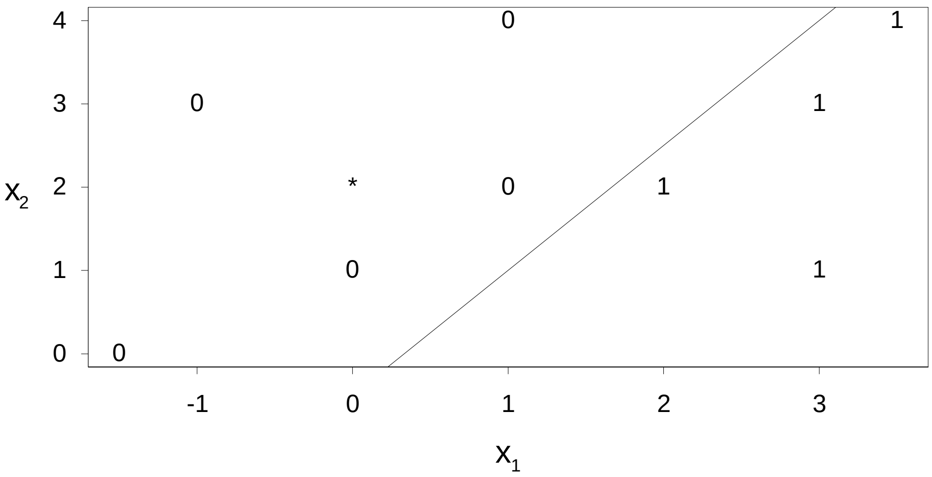

Fig. 1. Top view of a data set with explanatory variables x_1 and x_2 where $y_i = 0$ or $y_i = 1$. If the point indicated by the asterisk has $y_i = 0$ then the straight line completely separates the successes and failures.

Figure 2 plots the response y_i versus the linear combination $\mathbf{x}_i \mathbf{u}'$ where $\mathbf{u} \in \mathbb{R}^{p-1}$ is some direction. If the point denoted by $*$ in Figure 1 has $y_2 = 1$ it yields the point A in Figure 2. Then there is overlap, and the dashed line shows the MLE fit of the logistic regression. But if the point $*$ has $y_2 = 0$ we obtain the point B instead of A, and then there is complete separation. In that case the MLE estimate of θ does not exist. The dotted line shows the fitted curve after stopping an iterative MLE algorithm due to non-convergence. The vertical line separates the 0's and 1's in this plot. Of course, for higher-dimensional $\mathbf{x}_i$ it becomes harder to determine the overlap, and we will construct an algorithm to do so.

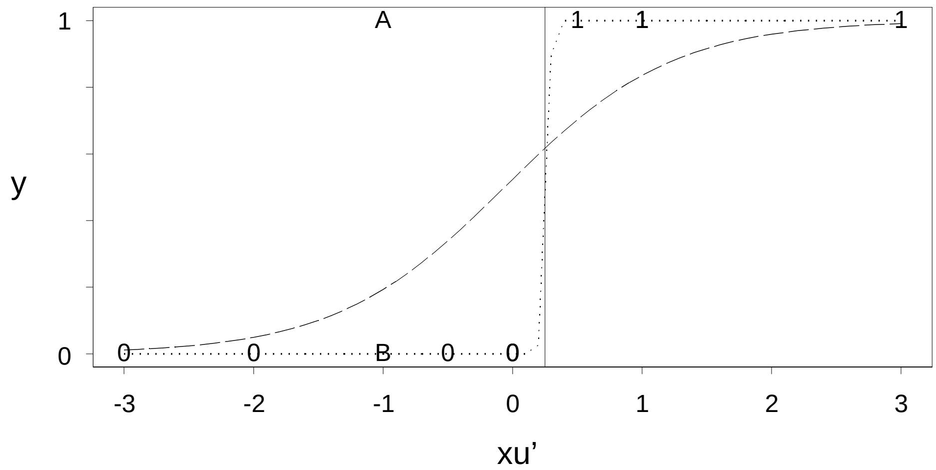

Fig. 2. Side view of the data set in Fig. 1 according to some direction $\mathbf{u}$. The asterisk in Fig. 1 corresponds either to point A or to point B.

3 Computing the overlap

There exists a simple connection between regression depth and complete separation. For a data set $Z_n = \{(x_{i,1}, \ldots, x_{i,p-1}, y_i);\ i = 1, \ldots, n\}$ with binary y_i we can consider the horizontal hyperplane given by $\theta^* = (0, \ldots, 0, 0.5)$. Then θ^* is a nonfit iff $n_{\text{complete}} = 0$, and more generally $n_{\text{complete}} = \text{rdepth}(\theta^*, Z_n)$. This implies that n_{complete} can be computed with an algorithm for the regression depth of a given hyperplane. For $p = 2$ the latter can be computed by the $O(n \log(n))$ time algorithm of Rousseeuw and Hubert (1999). For $p \geq 3$, Rousseeuw and Struyf (1998) constructed a fast approximation algorithm for the regression depth.

For n_{overlap} we cannot use the regression depth algorithms as they are, but we have constructed analogous algorithms for this case.

Some modifications of these algorithms can substantially reduce the computation time for data sets with a large number of ties, which is a common situation for binary regression models. Our algorithm consists of the following major steps.

1. Read the data set $Z_n = \{(x_{i,1}, \dots, x_{i,p-1}, y_i);\ i = 1, \dots, n\} \subset \mathrm{I\!R}^p$, where $y_i \in \{0, 1\}$, $i = 1, \dots, n$. Standardize the $x-$variables.

2. Determine the number of *different* points $(x^a_{j,1}, \dots, x^a_{j,p-1}, y^a_j)$ in Z_n, say n_a. For each $j \in \{1, \dots, n_a\}$ count the number t_j of tied data points, hence $n = \sum_{j=1}^{n_a} t_j$. From now on we will work with the aggregated data set $Z^a_n = \{(x^a_{j,1}, \dots, x^a_{j,p-1}, y^a_j; t_j); 1 \le j \le n_a\}$.

3. If $p = 2$, apply the exact algorithm for n_{overlap} (or n_{complete}) to the aggregated data set. Go to Step 7.

4. If $p > 2$, use the approximation algorithm based on projections. Define the number NITER of subsamples to be drawn. Initialize the random number generator. Set NSIN=0, ITER=1, and $n_{\text{overlap}} = n$ (or $n_{\text{complete}} = n$).

5. Draw a random subsample of size $p - 1$ from Z^a_n. If the $\{(x^a_{j,1}, 1)', \dots, (x^a_{j,p-1}, 1)'\}$ are linearly dependent (i.e. not of full column rank), set NSIN=NSIN+1 and draw the next random subsample. Else go to Step 6.

6. Project all $\mathbf{x}^a_j$ on the direction $\mathbf{u}$ orthogonal to the hyperplane given by the subsample. Aggregate the two-dimensional data set $\{(\mathbf{x}^a_j\mathbf{u}', y_j); j = 1, \dots, n_a\}$ and the corresponding counts t_j as defined in Step 2 and count the ties. Compute the two-dimensional n_{overlap}. If it is less than the current value of n_{overlap}, update the latter (or the same with n_{complete}). Set ITER=ITER+1. If ITER > NITER go to Step 7, else go to Step 5.

7. Output the resulting n_{overlap} (or n_{complete}), the corresponding direction $\mathbf{u}$, and for $p > 2$ the number NSIN of singular subsamples that were encountered.

The actual implementation is available from the first author at `A.Christmann@hrz.uni-dortmund.de` .

4 Examples

In this section we consider some data sets commonly used as test data in logistic regression. The values of n_{complete} and n_{overlap} were computed by the algorithm of Section 3, and are given in Tables 1 and 2. These tables also list the indices of important cases whose deletion would destroy the overlap, the computing times (on a Pentium PC with 166 MHz), and the trial number of the first occurrence of the final value of n_{complete} or n_{overlap}. We checked our results by first trying 10^4 subsamples and then 10^5 subsamples. For the data sets considered here, the final result was obtained already for 10^4 subsamples. The effort to compute n_{complete} or n_{overlap} was small to moderate, with computation times ranging between 2 seconds and 6 minutes. The computation time increases approximately linearly with the number of subsamples being drawn.

Finney (1947) lists the vaso constriction data set about a controlled experiment to study the effect of the rate and volume of air on a transcient reflex vaso-constriction in the skin of the digits. Pregibon (1981) uses this data set to illustrate his diagnostic measures for detecting outlying observations and quantifying their impact on various aspects of the maximum likelihood fit. We use this data set with the same explanatory variables log(rate) and log(volume). Pregibon (1981) shows that cases 4 and 18 are outlying and that both have a large impact on the MLE fit. Both cases are downweighted by the M-estimators in Künsch, Stefanski and Carroll (1989, p. 465). We find $n_{\text{overlap}} = n_{\text{complete}} = 3$ in Tables 1 and 2. The well-known outliers 4 and

18 stick out in Figure 3a.

Table 1: The number n_{complete} for several data sets.

Data set (n,p)	n_{complete}	important cases	number of samples 10,000	100,000	trial number
Vaso constriction $(39,3)$	3	4,18,29 or 4,18,24	2 sec	23 sec	36
Cancer remission $(27,7)$	3	7,23,24	8 sec	86 sec	2472
Food stamp $(150,4)$	17	5,22,40,44,51,66, 73,79,95,103,106, 109,113,120,135, 137,147	8 sec	89 sec	223
IVC $(3200,5)$	458	not given	8 sec	75 sec	7705
Hemophilia $(52,3)$	0	–	3 sec	31 sec	9
Birth weight $(189,11)$	47	not given	37 sec	371 sec	5253

Table 2: The number n_{overlap} for several data sets.

Data set (n,p)	n_{overlap}	important cases	number of samples 10,000	100,000	trial number
Vaso constriction $(39,3)$	3	4,18,29 or 4,18,24	2 sec	23 sec	36
Cancer remission $(27,7)$	3	7,23,24 or 2,8,15	8 sec	86 sec	274
Food stamp $(150,4)$	6	22,66,103, 120,137,147	9 sec	88 sec	5
IVC $(3200,5)$	213	$\{x_{i,4} = 1 \text{ and } y_i = 0\}$	7 sec	75 sec	20
Hemophilia $(52,3)$	0	–	3 sec	31 sec	9
Birth weight $(189,11)$	5	13,51,93,102,106	37 sec	374 sec	4

The cancer remission data set taken from Lee (1974) consists of patient characteristics. Cancer remission is the response variable. We find $n_{\text{overlap}} = n_{\text{complete}} = 3$. Case 24 seems to be somewhat extreme in Figure 3b.

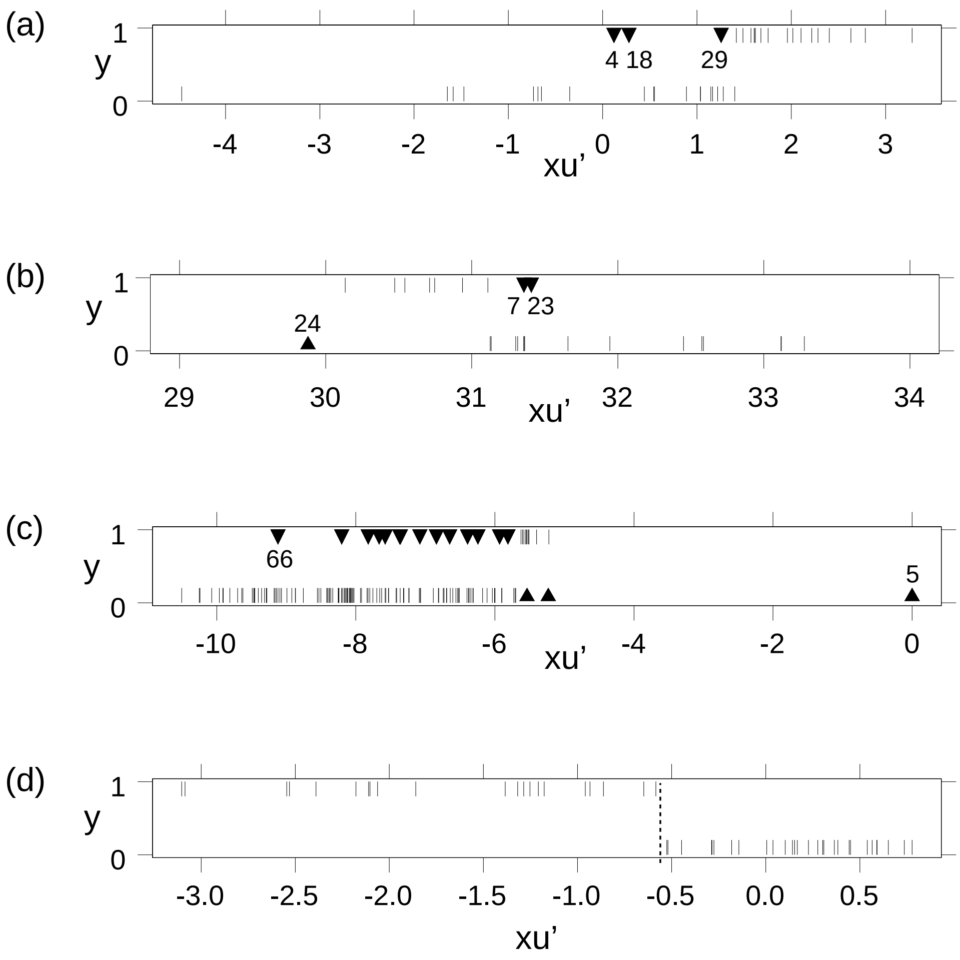

Figure 3: Plot of y_i versus $\mathbf{x}_i\mathbf{u}'$ with $\mathbf{u}$ yielding the smallest n_{complete} for (a) the vaso constriction data; (b) the cancer remission data; (c) the food stamp data; and (d) the hemophilia data. If one would remove the points marked as triangles, the data would be completely separated.

Künsch, Stefanski and Carroll (1989) and Carroll and Pederson (1993) investigate the food stamp data set using M-estimators. Some observations were strongly downweighted. Case 5 is isolated in the design space, and appears to be an outlier in the $y-$direction. Case 66 is somewhat outlying too. Künsch, Stefanski and Carroll (1989) concluded that " ... it should be checked how close the data are to indeterminacy". For the food stamp data set we find $n_{\text{overlap}} = 6$ and $n_{\text{complete}} = 17$. Our approach draws the data analyst's attention to the same two cases 5 and 66 in Figure 3c.

Jaeger et al. (1997, 1998) carry out an in vitro experiment to study possible risk factors of the thrombus-capturing efficacy of inferior vena cava (IVC) filters. We focus on the study of a particular conical IVC filter, for which the design consisted of 48 different vectors of the form $(x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4})$. For each vector there were m_i replications with $m_i \in \{50, 60, 90, 100\}$, yielding a total of $n = 3200$. The IVC data set is listed in Table 3 in aggregated form. The explanatory variables are: thrombus diameter $x_{i,1}$ (continuous, 1.5mm to 8.5mm), inferior vena cava diameter (discrete; $[x_{i,2} = 0, x_{i,3} = 0]$ if 20mm; $[x_{i,2} = 1, x_{i,3} = 0]$ if 24mm; $[x_{i,2} = 0, x_{i,3} = 1]$ if 28mm), and thrombus length (discrete; $x_{i,4} = 0$ if short; $x_{i,4} = 1$ if long). The IVC data set has many ties. We find $n_{\text{complete}} = 458$. It is interesting to note that there is no overlap if the $n_{\text{overlap}} = 213$ cases with $x_{i,4} = 1$ and $y_i = 0$ are dropped, where long thrombi were investigated and failures were observed.

Hermans and Habbema (1975) investigate a data set with 30 women known to be non-carriers of hemophilia and 22 women who are carriers of hemophilia. There are two continuous explanatory variables. The data set is completely separated and we find $n_{\text{complete}} = n_{\text{overlap}} = 0$ in Figure 3d.

Hosmer and Lemeshow (1989) give a data set on 189 births at a US hospital. There are 10 explanatory variables, and low birth weight is used as the binary response variable. Our algorithms find $n_{\text{complete}} = 47$ and $n_{\text{overlap}} = 5$, the latter being surprisingly low.

Table 3: Inferior vena cava (IVC) data set, where m_j is the number of replications in each design point, with $\sum y_j$ successes and $m_j - \sum y_j$ failures.

$x_{j,1}$	$x_{j,2}$	$x_{j,3}$	$x_{j,4}$	$\sum y_j$	m_j	$x_{j,1}$	$x_{j,2}$	$x_{j,3}$	$x_{j,4}$	$\sum y_j$	m_j
1.5	0	0	0	16	90	5.5	0	0	0	84	90
1.5	0	0	1	74	100	5.5	0	0	1	97	100
1.5	1	0	0	5	50	5.5	1	0	0	37	50
1.5	1	0	1	36	50	5.5	1	0	1	41	50
1.5	0	1	0	3	50	5.5	0	1	0	40	50
1.5	0	1	1	30	60	5.5	0	1	1	40	60
2.5	0	0	0	24	90	6.5	0	0	0	89	90
2.5	0	0	1	95	100	6.5	0	0	1	98	100
2.5	1	0	0	4	50	6.5	1	0	0	48	50
2.5	1	0	1	38	50	6.5	1	0	1	42	50
2.5	0	1	0	5	50	6.5	0	1	0	40	50
2.5	0	1	1	51	60	6.5	0	1	1	51	60
3.5	0	0	0	52	90	7.5	0	0	0	89	90
3.5	0	0	1	95	100	7.5	0	0	1	97	100
3.5	1	0	0	18	50	7.5	1	0	0	49	50
3.5	1	0	1	42	50	7.5	1	0	1	49	50
3.5	0	1	0	25	50	7.5	0	1	0	47	50
3.5	0	1	1	52	60	7.5	0	1	1	53	60
4.5	0	0	0	80	90	8.5	0	0	0	90	90
4.5	0	0	1	95	100	8.5	0	0	1	99	100
4.5	1	0	0	23	50	8.5	1	0	0	48	50
4.5	1	0	1	38	50	8.5	1	0	1	49	50
4.5	0	1	0	22	50	8.5	0	1	0	47	50
4.5	0	1	1	46	60	8.5	0	1	1	59	60

5 Summary

There is an interesting relation between the notion of regression depth introduced by Rousseeuw and Hubert (1999) and the notion of separation developed by Albert and Anderson (1984) and Santner and Duffy (1986). The latter authors investigate conditions under which the maximum likelihood estimate of θ exists, in a logistic regression model with an intercept term. They showed that if the data set is completely or quasicompletely separated, then the maximum likelihood estimate of θ does not exist. If the data set has overlap, then the maximum likelihood estimate of θ exists and it is unique. In the present paper algorithms are proposed to determine n_{overlap}, the smallest number of observations whose removal would destroy the overlap. In our terminology, having overlap means that $n_{\text{overlap}} > 0$. The examples in Table 2 illustrate that n_{overlap} is often quite small, especially in higher dimensions, so that the result of a logistic regression often depends crucially on only a few observations.

If the assumptions of a logistic regression model for binary response variables are valid, it holds for any parameter vector $\theta \in \mathbb{R}^p$ that

$$\begin{aligned} \mathrm{P}_\theta(n_{\text{complete}} = 0) &\geq \mathrm{P}_\theta(\text{all}\, Y_i = 0) + \mathrm{P}_\theta(\text{all}\, Y_i = 1) \\ &= \prod_{i=1}^{n} [1 - \Lambda((\mathbf{x}, 1)\theta')] + \prod_{i=1}^{n} \Lambda((\mathbf{x}, 1)\theta') > 0\,. \end{aligned}$$

This is why there are no estimators that always have a high finite-sample (replacement) breakdown value in the sense of Donoho and Huber (1983) for logistic regression with binary response variables, c.f. Christmann (1994).

The algorithm for n_{overlap} is also useful in other regression models with binary response variables. For instance, the probit model uses the cumulative distribution function Φ of the standard normal distribution instead of Λ in (1), and n_{overlap} has the same importance as in logistic regression.

Acknowledgements

The authors thank Prof. R.J. Carroll for making available the Food Stamp data set, and Dr. H.J. Jaeger for making available the IVC data set in Table 3.

References

Albert, A. and Anderson, J.A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1-10.

Carroll, R.J. and Pederson, S. (1993). On robust estimation in the logistic regression model. *J. R. Statist. Soc.* B. **55**, 693-706.

Christmann, A. (1994). Least median of weighted squares in logistic regression with large strata. *Biometrika* **81**, 413-417.

Cox, D.R. and Snell, E.J. (1989). *Analysis of Binary Data, 2nd Edition.* Chapman and Hall, London.

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*, eds. P.J. Bickel, K.A. Doksum, and J.L. Hodges, Jr. Belmont, California: Wadsworth, 157-184.

Finney, D.J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34**, 320-334.

Jaeger, H.J., Mair, T., Geller, M., Kinne, R.K., Christmann, A., Mathias, K.D. (1997). A physiologic in vitro model of the inferior vena cava with a computer-controlled flow system for testing of inferior vena cava filters. *Investigative Radiology* **32**, 511-522.

Jaeger, H.J., Kolb, S., Mair, T., Geller, M., Christmann, A., Kinne, R.K., Mathias, K.D. (1998). In vitro model for the evaluation of inferior vena cava filters: effect of experimental parameters on thrombus-capturing efficacy of the Vena Tech-LGM Filter. *Journal of Vascular and Interventional Radiology* **9**, 295-304.

Hermans, J. and Habbema, J.D.F. (1975). Comparison of five methods to estimate posterior probabilities. *EDV in Medizin und Biologie* **6**, 14-19.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression.* Wiley, New York.

Künsch, H.R., Stefanski, L.A. and Carroll, R.J. (1989). Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *J. Amer. Statist. Assoc.* **84**, 460-466.

Lee, E.T. (1974). A computer program for linear logistic regression analysis. *Computer Programs in Biomedicine*, 80-92.

Pregibon, D. (1981). Logistic regression diagnostics. *Ann. Statist.* **9**, 705-724.

Rousseeuw, P.J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79**, 871-880.

Rousseeuw, P.J. and Hubert, M. (1999). *Regression Depth.* To appear in *J. Amer. Statist. Assoc.*, June 1999.

Rousseeuw, P.J. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing* **8**, 193-203.

Santner, T.J. and Duffy, D.E. (1986). A note on A. Albert and J.A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **73**, 755-758.