Validation of Rating Models

A. Miemiec[a]

[a] *FRAME Consulting GmbH, Gabriel-Max-Strasse 12, D-10245 Berlin, Germany*

Abstract

This article reviews score functions, important methodological building blocks of rating models. By emphasising the relationship to a Bayesian classification problem, different approaches to determine a score function are related to each other. This makes it possible to generate estimators for the score function with purely geometric means on a possibly expanded data basis, which in turn can also be used for the validation of a rating model.

Keywords: rating, methodology, validation

1. Introduction

For the purpose of this article, a rating model ($\mathcal{R}$) is a mapping that based on a tuple of d observable characteristics, x, assigns to a borrower for a given time horizon (e.g. 1 year) a probability of default, $PD(x) \in [0,1]$:

$$\mathcal{R}\left[Borrower\right] : \mathbb{R}^d \mapsto [0,1]. \tag{1}$$

This mapping belongs to a Bayesian classification problem, i.e. the assignment of a borrower to one of two groups $g \in \{I, II\}$. Group I represents the set of defaulted borrowers and group II the set of non-defaulted borrowers.

If the characteristics, x, assigned to a borrower are realisations of a vector of continuous random variables, X, and if the group indicator, g, is a realisation of a discrete random number, G, with the possible values $\{I, II\}$ then the group assignment is done according to the following rule:

$$g(x) \quad = \quad \underset{g \in \{I,II\}}{\operatorname{argmax}} \, \mathbb{P}(G = g \,|\, X = x). \tag{2}$$

The general practice in credit risk is to use score functions to actually perform the classification. Therefore, the following methodological question arises: What is the relationship between a score function and this classification rule?

In the following two cases must be distinguished. Either the classification according to eq.(2) can be done by analytic means or it requires the application

Email address: andre.miemiec@frame-consult.de (A. Miemiec)

of a regression procedure.

Case A: In a convenient situation, where the conditional densities of the characteristics x, given that the observations belong to the group g, $\rho_X(x|G=g)$, are multivariate normal densities and the prior probabilities of default of the groups, $\{\pi_k\}_{k=I,II}$, are known, the probability of belonging to the group $G=g$ given $X=x$ is calculated from eq.(A.1) and eq.(A.2) in the appendix and reads

$$\mathbb{P}(G=g|X=x) \;=\; \frac{e^{-\tilde{d}_g^2(x)/2}}{\sum_i e^{-\tilde{d}_i^2(x)/2}}.$$

The distance $\tilde{d}_g^2(x)$ is defined in eq.(A.3). A maximum probability corresponds to a minimum distance. The classification can then also be done via a score function. In the binary case, the score function takes the following form:

$$Score(x) \;=\; \frac{1}{2}\left(\tilde{d}_{II}^2(x) - \tilde{d}_I^2(x)\right). \tag{3}$$

According to the above definitions, a score greater than zero leads to an assignment of a borrower to group I and a score less than zero leads to an assignment of a borrower to group II. The decision boundary is defined by the set of points ξ for which the new distances are identical:

$$\tilde{d}_I^2(\xi) \quad = \quad \tilde{d}_{II}^2(\xi).$$

The decision boundary turns out to be a plane if the two modified Mahalanobis distances in the score of eq.(3) are employing the same covariance matrices ($\Sigma_I = \Sigma_{II}$). In this particular situation, the coefficients of the characteristics in the score function determine the normal direction of the decision boundary. The absolute position of the decision boundary along the axis, given by the normal direction of the decision boundary, is determined by the prior probabilities, i.e. the PD-profile.

Case B: In the situation, where either the conditional densities of the characteristics x, given that the observations belong to the group g, or the prior probabilities of default of the groups are unknown[1], the decision according to eq.(2) is done as follows: For the problem with 2 groups, we use

$$\mathbb{P}(G=g|X=x) = \mathbb{E}\left[\mathbb{1}_g|X=x\right].$$

The transformation of the conditional probability into a conditional expected value allows the use of regression methods, where the regression function is in turn associated with a linear score. This regression model can be calibrated to

[1] The under-determined nature of the prior probabilities of the reference group of borrowers follows from the fact, that the PD-profile of a bank portfolio will usually not correspond to the PD-profile of the whole population (e.g. due to a special business strategy of the bank).

data specific for a given bank. In the binary case, the regression can be carried out, for example, with the help of a logit model (cf. section Appendix A.3):

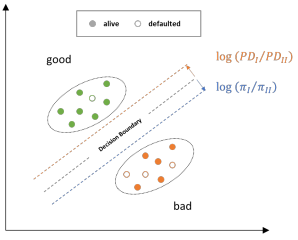
$$\mathbb{P}(G = I \,|X = x) \quad = \quad \frac{1}{1 + \exp[\,-(\mathfrak{a} + \langle\, \mathfrak{b},\, x \,\rangle)\,]}.$$

Here the constant $\mathfrak{a}$ and the constant vector $\mathfrak{b}$ are the parameters of a multi-linear regression function.

2. Application

We do heuristics. For this purpose we stay within the framework of a binary model. We assume that the default-weighted and group-specific distributions of the characteristics follow the same multivariate normal distributions as if no conditioning on the default event were performed. This is by no means implausible, since a classification into good and bad borrowers (e.g. based on an expert judgement) should more or less correspond to a classification with respect to a default flag. Otherwise, the expert was not well chosen.

Then, in principle, the score function of a logistic regression model can be constructed almost completely geometrically. If one assumes for the sake of simplicity that the covariances of the two groups match, one proceeds as follows: One starts with the calculation of the parameters of the multivariate normal distribution of the ensemble of characteristics x. This results in a geometric score according to eq.(3), which is initially calibrated to arbitrarily chosen prior probabilities, π_i, (blue dashed line on the left hand side of Fig.1). However, if one is interested in the mapping $\mathcal{R}$ of eq.(1) for a specific bank, the very same score has to be corrected for the bank-specific prior default probabilities, PD_i, (red dashed line on the left hand side of Fig.1). Here, the two score functions differ only by constants that are related to specific choices of prior probabilities. The other parameters are already geometrically determined and represent in the given situation the normal direction of the plane, which is the decision boundary.



	Stylised Example
Parameters	$\mu_1 = (+1,+1,+1,+1,+1), \quad \mu_2 = (-1,-1,-1,-1,-1)$ $\Sigma_1 = \Sigma_{II} = 1_{5x5}$ $\pi_I = \pi_{II} = 0.5$
Score	Geometric Approach (cf. eq. (3)): Score = $2.0 \cdot X_1 + 2.0 \cdot X_2 + 2.0 \cdot X_3 + 2.0 \cdot X_4 + 2.0 \cdot X_5$ LDA (numerical): Score = $2.1 \cdot X_1 + 0.9 \cdot X_2 + 1.8 \cdot X_3 + 0.9 \cdot X_4 + 1.5 \cdot X_5$

Figure 1: Geometric estimation of the score function of a rating model (Left: schematic sketch, Right: stylised example)

The right hand side of Fig.1 displays the results of a stylised example, where the score function is estimated once by the geometric approach and once by a

numerical implementation of the linear discriminant analysis (LDA). The sample consists of 60 observations of five characteristics and a default flag. The in-sample priors of belonging to the groups of defaulted and non-defaulted observations are 50% each. Apart from a normalisation of the scale, the two scores essentially agree on the normal direction of the decision boundary. Any differences between the two scores are due to numerical issues caused by sparse number of observations.

3. Summary

This article described a hands on approach to the validation of a rating model. The added value is the effective construction of a geometric and functionally transparent score function (proxy) that can be used to assess the quality of the outcome of a regression model. If necessary, the geometric score function can be constructed using an expanded data basis. The essential parts of the geometric score, that correspond to the normal direction of the decision boundary, can then be calculated by circumventing the explicit knowledge of default events. The result is suitably adjusted to a given profile of default probabilities.

The geometric score is only an estimator for the regression function. From a theoretical point of view, this has at least two reasons: The first reason is the specific choice of a maximum likelihood function, which is needed to determine the regression model. A second reason is the possible uncertainty about the prior probabilities.

The validation of of a regression model as proposed here covers at least two important use cases: First, the frequently asked question of whether the relative weights of the characteristics within a score function are plausible can be answered by geometric means. Second, stability analyses for a rating model can be carried out with the same methodology. By comparing the geometric decision boundaries on the basis of bank-specific data and on the basis of (exogenous) data, that represent the entire population, bank-specific properties (e.g. weaknesses) of a rating model can be systematically identified and assessed.

Acknowledgments: The author would like to thank Tilman Wolff-Siemssen and another unnamed person for useful comments on earlier versions of this article. All remaining errors are the author's responsibility.

Appendix A. Formulas

Appendix A.1. Bayesian Identity

We assume that the characteristics, x, assigned to a borrower are realisations of a vector of continuous random variables, X, and the group indicator, g, is a realisation of a discrete random number, G, with possible values $\{I, II\}$.

Let $\mathbb{P}(G = g) = \pi_g$ be the probability that a randomly selected observation belongs to the group g, and $\rho_X(x|G = g)$ be the conditional probability density

function of the random vector X, given that the observation belongs to the group g.

We are interested in the conditional probability, $\mathbb{P}(G = g|X = x)$, that an observation is from the group g, given the observed values x of the random vector X. Using the usual notation of Bayes' rule,

$$\mathbb{P}(B|A) = \mathbb{P}(A \text{ and } B)/\mathbb{P}(A),$$

where event A is the observation of characteristics x and event B is the observation of membership of group g, the conditional probability sought can be formally calculated as follows:

$$\mathbb{P}(G = g|X = x) \;=\; \frac{\mathbb{P}(X = x \text{ and } G = g)}{\mathbb{P}(X = x)}.$$

The numerator of the above expression is the probability weight that a randomly selected observation both has the characteristics x and is from the group g. This is $\rho_X(x|G = g) \cdot \pi_g$. The denominator is the unconditional probability weight that a randomly selected observation has the characteristics x across all groups. This is

$$\mathbb{P}(X = x) = \textstyle\sum_i \rho_X(x|G = i) \cdot \pi_i.$$

By composing and rearranging the above formulae yields the identity

$$\mathbb{P}(G = g|X = x) \quad = \quad \frac{\rho_X(x|G = g) \cdot \pi_g}{\sum_i \rho_X(x|G = i) \cdot \pi_i}. \tag{A.1}$$

Appendix A.2. Modified Mahalanobis Distance

If $f_X(x; \mu_i, \Sigma_i)$ denotes for a vector of multivariate normally distributed random numbers, X, the density of the d-dimensional normal distribution with mean μ_i and covariance Σ_i and π_i the prior probability of belonging to the group i, then

$$f_X(x; \mu_i, \Sigma_i) \cdot \pi_i \quad = \quad \frac{1}{(2\pi)^{d/2}}\, e^{-\tilde{d}_i^2(x)/2} \tag{A.2}$$

with

$$\tilde{d}_i^2(x) = (x - \mu_i)^T \cdot \Sigma_i^{-1} \cdot (x - \mu_i) + \log\det\Sigma_i - 2\log\pi_i. \tag{A.3}$$

Appendix A.3. Logistic Function

For the analytically tractable case A in section 1, the score function can be represented as a quotient of conditional probabilities. In the binary case, for example, the well-known log-odds read:

$$\log\frac{\mathbb{P}(G = I\,|X = x)}{\mathbb{P}(G = II|X = x)} \quad = \quad Score(x). \tag{A.4}$$

A little algebra then leads to a logit type representation:

$$\mathbb{P}(G = I \,|X = x) \quad = \quad \frac{1}{1 + \exp[\,-Score(x)\,]}$$

For the case of linear scores, this is the logistic function.

The more general case B from section 1 can also rely on this ansatz. However, in this situation the explicit assumption of 'linear log-odds' must be made, i.e. the score analogous to the score in eq.(A.4) is assumed to be a linear function.