Logistic Regression

MATH 271: Statistical Methods

First Semester, S.Y. 2021-2022

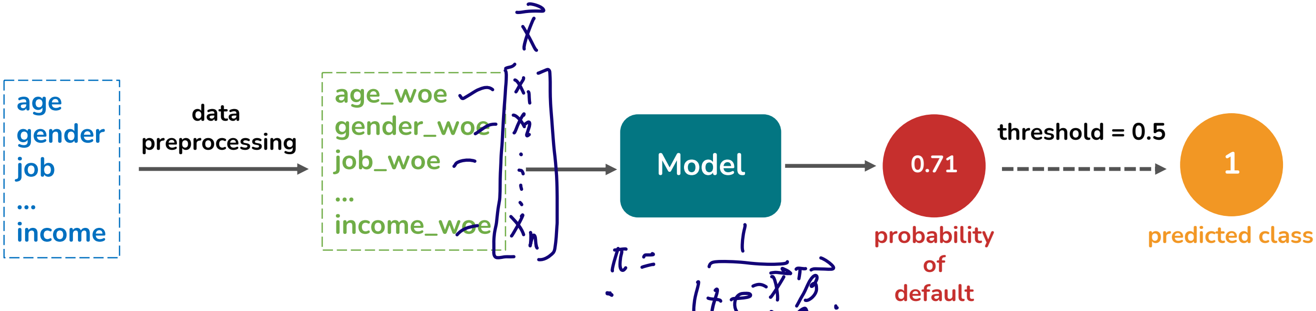
Ateneo de Manila University

This session is recorded.

Binary Classification for Credit Scorecards

class = 1 if the account will default (bad account)

class = 0 if the account will not default (good account)



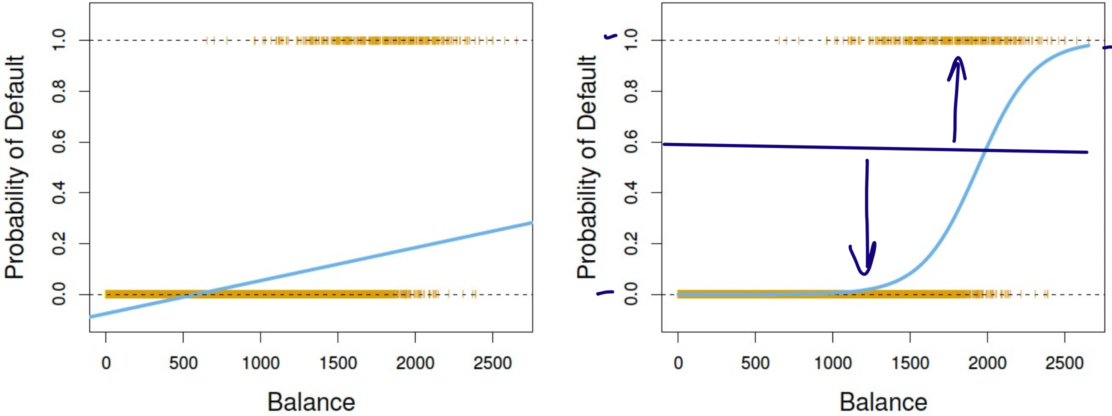
$\pi = \frac{1}{1+e^{-\vec{X}^T\vec{\beta}}}$.

$$\text{predicted class} = \begin{cases} 1, & p \geq \text{threshold} \\ 0, & p < \text{threshold} \end{cases}$$

how to find $\vec{\beta}$?

MLE

Linear Regression vs Logistic Regression



→ logistic curve

$$y = \frac{1}{1+e^{-x}}$$

$$y_i = \begin{cases} 1 & \text{w/ prob. } \pi_i \\ 0 & \text{w/ prob. } 1-\pi_i \end{cases}$$

class for the ith account

Distribution of Y: $Y_i \sim Be(\pi_i)$

prob. of default for the ith account

- Y_i has a probability density function of

$$f(Y_i) = \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}$$

$$Y_i = \begin{cases} 1, & \text{with probability } \pi_i \, . \\ 0. & \text{with probability } 1-\pi_i \, . \end{cases}$$

- The mean and variance of Y_i is given by

$$0 \le \mathbb{E}(Y_i) = \underline{\pi_i} \le 1$$

$$\mathrm{Var}(Y_i) = \pi_i(1-\pi_i)$$

- Each of the $Y_i{}'s$ are independent of each other

Definition: Link Function

The **link function** of the dependent variable Y is the transformation on Y that provides the linear relationship between the linear predictor, $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}$ and $\mathbb{E}(Y)$. That is, $g(\cdot)$ is a link function for the dependent variable Y if

$$g\big(\mathbb{E}(Y)\big) = \underline{\underline{\boldsymbol{X}}}^{\boldsymbol{T}}\boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

$$g(\underline{\underline{\pi}})$$

Remark: $\boldsymbol{X} = [1, X_1, X_2, \ldots, X_k]^T$ *and* $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_k]^T$

Definition: Logit Link Function

logistic curve

The **logit link function** gives the log-odds that $Y = 1$. This means that

$$g(\pi) = \vec{X}^T\vec{\beta}$$

$$\mathrm{logit}(\mathbb{E}(Y)) = \boxed{\ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X}^T\boldsymbol{\beta}} \quad ✓$$

$$g(x) = \ln\left(\frac{x}{1-x}\right)$$

The probability π can then be computed as

$$\pi = \frac{1}{1+\exp(-\boldsymbol{X}^T\boldsymbol{\beta})}$$

✓

$$\pi = \frac{1}{1+e^{-X^T\beta}} \longrightarrow \text{logistic curve}$$

$$\ln\left(\frac{\pi}{1-\pi}\right) = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}$$

$$\frac{\pi}{1-\pi} = \exp\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}\right)$$

$$\pi = \exp\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}\right)(1-\pi)$$

$$\pi = \exp\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}\right) - \pi\exp\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}\right)$$

$$\pi\left(1+\exp\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}\right)\right) = \exp(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta})$$

$$\pi = \frac{\exp\left(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta}\right)}{1+\exp(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta})} = \frac{1}{1+\exp(-\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\beta})}$$

Model Parameters

We will be using the method of maximum likelihood to obtain the regression parameters $\beta_0, \beta_1, \ldots, \beta_k$

$\vec{\beta}$

Consider n independent Bernoulli observations (*referring to n accounts),* $Y_1, Y_2, \ldots, Y_n$. The probability distribution function of Y_i is given by

$$\boxed{f(Y_i) = \pi_i^{Y_i}(1 - \pi_i)^{1-Y_i}}$$

Model Parameters

Hence, we aim to choose $\boldsymbol{\beta}$ that maximizes the likelihood function

maximize the

$$L(\boldsymbol{\beta}) = \boxed{\prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}}$$

f(Y_i)

Or, the log-likelihood function

$$\ln(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(Y_i \ln(\pi_i) + (1-Y_i)\ln(1-\pi_i)\right)$$

maximize ln(L(β))

R

optimization problem ⇒ Lagrange multiplier

Probabilities of Default

With this, we should be able to get the parameter estimates $\hat{\boldsymbol{\beta}}$ and thus our model. The estimates of the probabilities π_i of default for account i can be obtained from an equation we had earlier

$$\hat{\pi}_i = \frac{1}{1 + \exp\left(-\boldsymbol{X}^{\boldsymbol{T}}\hat{\boldsymbol{\beta}}\right)} \quad ✓$$

Which can be used to determine probabilities for new data sets.

optimal $\vec{\beta}$ vector

What should the $X_i's$ be?

We want to avoid inconsistencies from units of variables.

prob. of default

$$\text{WOE} = \ln\left[\frac{P(c|Good)}{P(c|Bad)}\right]$$

Hence, our logistic regression model is given by

$$\underbrace{\ln\left(\frac{\pi_i}{1-\pi_i}\right)}_{\text{log odds that the account will default}} = \beta_0 + \beta_1 WOE_{1,i} + \beta_2 WOE_{2,i} + \cdots + \beta_k WOE_{k,i}$$

$= \vec{X}^T \beta$

In this model, $\underline{\beta_i < 0}$ for $i = 1,2,3, \dots, k$

Evaluation Metrics for Classification Tasks

MATH 271: Statistical Methods

First Semester, S.Y. 2021-2022

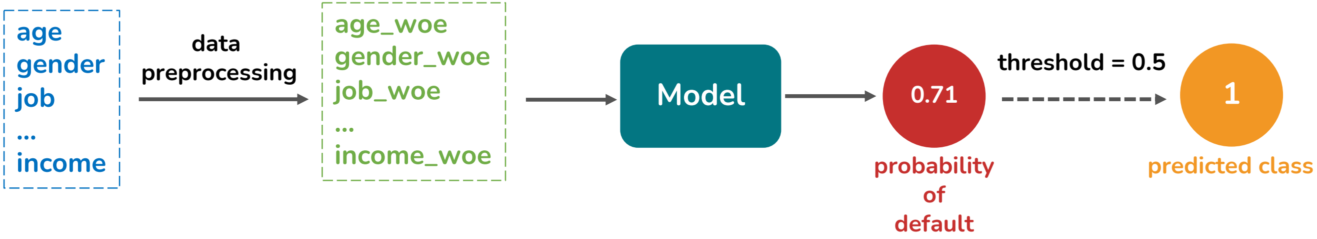
Ateneo de Manila University

This session is recorded.

Binary Classification for Credit Scorecards

class = 1 if the account will default (bad account)

class = 0 if the account will not default (good account)



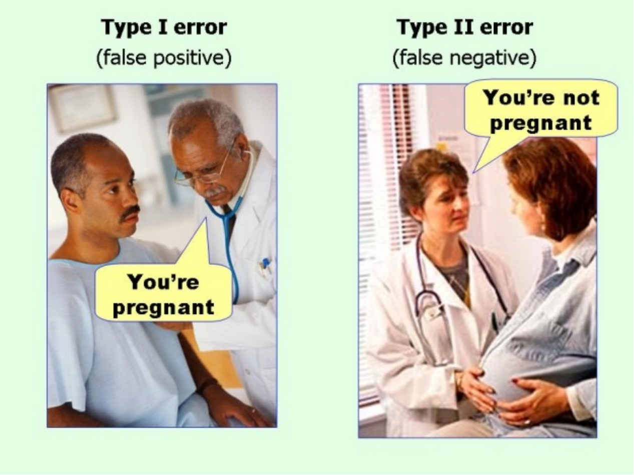
$$\text{predicted class} = \begin{cases} 1, & \text{p} \geq \text{threshold} \\ 0, & \text{p} < \text{threshold} \end{cases}$$

Classification

class = 1 if the account will default (bad account)

class = 0 if the account will not default (good account)

predicted class	actual class	
1	1	true positive (TP)
0	0	true negative (TN)
1	0	false positive (FP)
0	1	false negative (FN)



Example

class = 1 if the account will default (bad account)

class = 0 if the account will not default (good account)

Suppose out of 150 accounts in the dataset, we have the following **confusion matrix**:

Actual

		1	0
Predicted	1	40	30
	0	10	70

1. How many accounts are TP? TN? FP? FN?
2. What is the bad rate of the dataset?

Evaluation Metrics

1. Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN}$$

How many of the accounts were correctly classified by the model?

2. Precision

$$\frac{TP}{TP + FP}$$

How accurate is the model when it is trying to identify positive samples (bad accounts)?

Evaluation Metrics

3. Sensitivity (True Positive Rate)

$$\frac{TP}{TP + FN}$$

How many of the actual bad accounts were correctly classified by the model?

4. Specificity (True Negative Rate)

$$\frac{TN}{TN + FP}$$

How many of the actual good accounts were correctly identified by the model?

Evaluation Metrics

5. False Positive Rate

$$\frac{FP}{TN + FP}$$

How many of the actual good accounts were misclassified by the model?

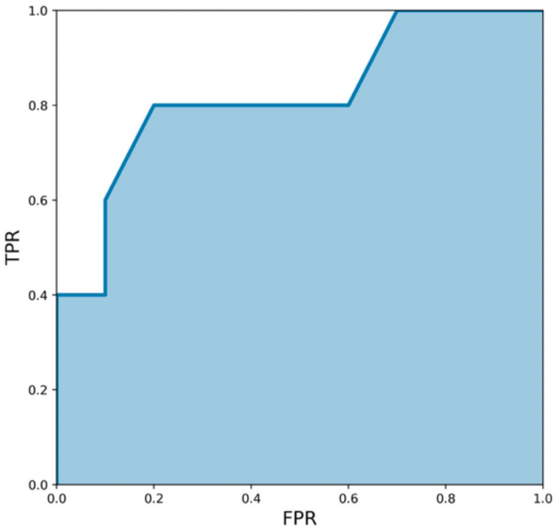
6. F1 score

$$\frac{2 * \text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

harmonic mean of precision and sensitivity

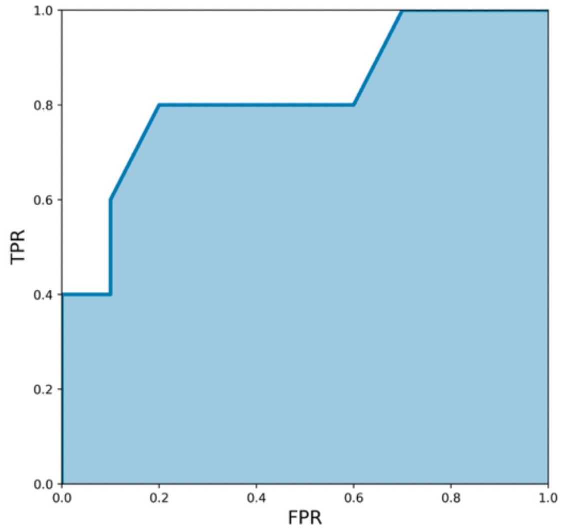
The value of threshold matters!

Threshold	TPR	FPR
0.0	1.0	1.0
0.1	1.0	0.9
0.2	1.0	0.7
0.3	0.8	0.6
0.4	0.8	0.3
0.5	0.8	0.3
0.6	0.8	0.2
0.7	0.6	0.1
0.8	0.6	0.1
0.9	0.4	0.0
1.0	0.0	0.0



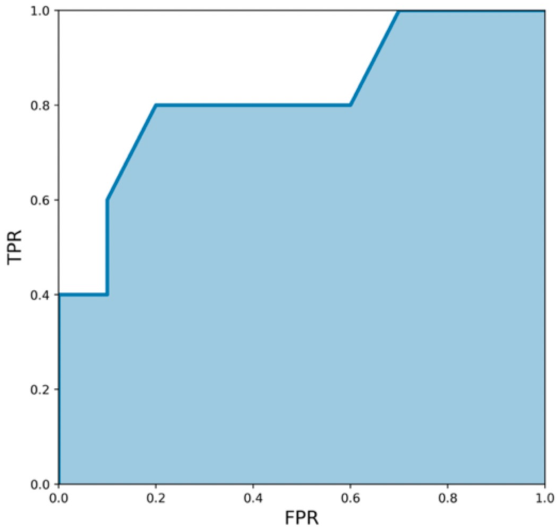
Receiver Operating Characteristic (ROC)

- This curve is also known as the **Receiver Operating Characteristic (ROC)**.
- The **Area Under ROC Curve** or **Area Under Curve (AUC)** is another metric which is frequently used for imbalanced datasets.
- There are many ways to calculate AUC, you may just use R/Python packages to compute for it.



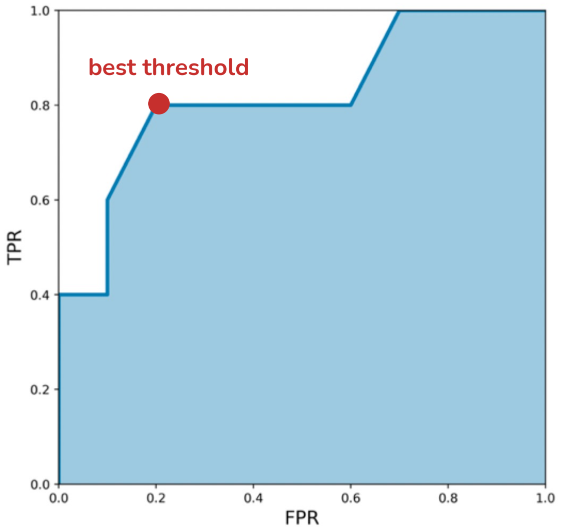
AUC values range from 0 to 1

- The higher the AUC, the better is the model’s performance on the dataset!
- AUC values below 0.5 imply that your model is worse than random.
- Ex: AUC = 0.85 implies that if you select a random bad account (positive sample) and another random good account (negative sample) from the dataset, then the model will choose the bad account over the good account as a positive class with a probability of 0.85.



You can use the ROC curve to find the best threshold!

Threshold	TPR	FPR
0.0	1.0	1.0
0.1	1.0	0.9
0.2	1.0	0.7
0.3	0.8	0.6
0.4	0.8	0.3
0.5	0.8	0.3
0.6	0.8	0.2
0.7	0.6	0.1
0.8	0.6	0.1
0.9	0.4	0.0
1.0	0.0	0.0



Other evaluation metrics

1. Somer’s D
2. Kolmogorov-Smirnov (KS) Statistic
    1. You can use this as an alternative to find the optimal threshold (cut-off probability)
3. Cumulative Accuracy Profile (CAP) and Gini Coefficient

Goodness-of-Fit Statistical Tests

1. KS Test
2. Hosmer-Lemeshow Test

Score Scaling

MATH 271: Statistical Methods

First Semester, S.Y. 2021-2022

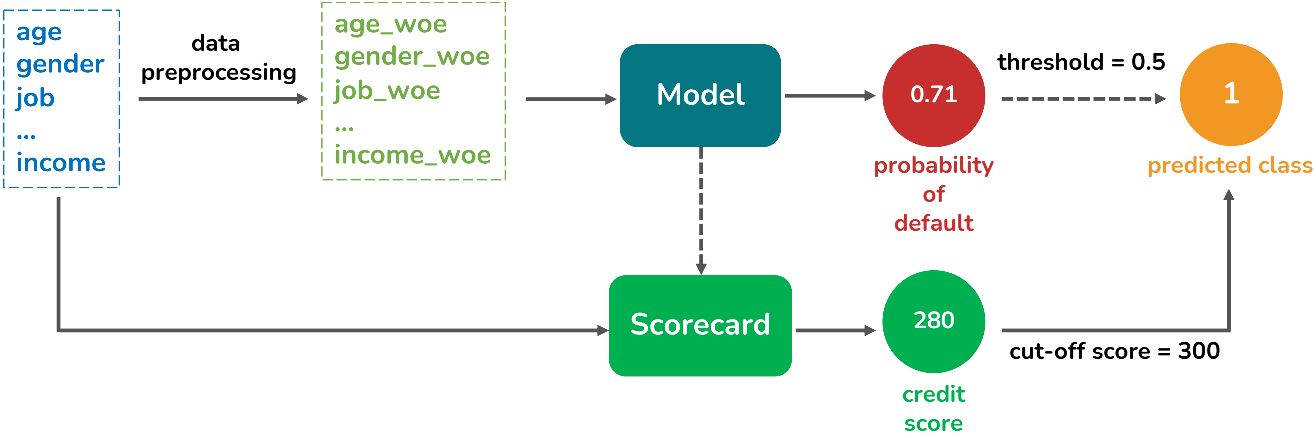
Ateneo de Manila University

This session is recorded.

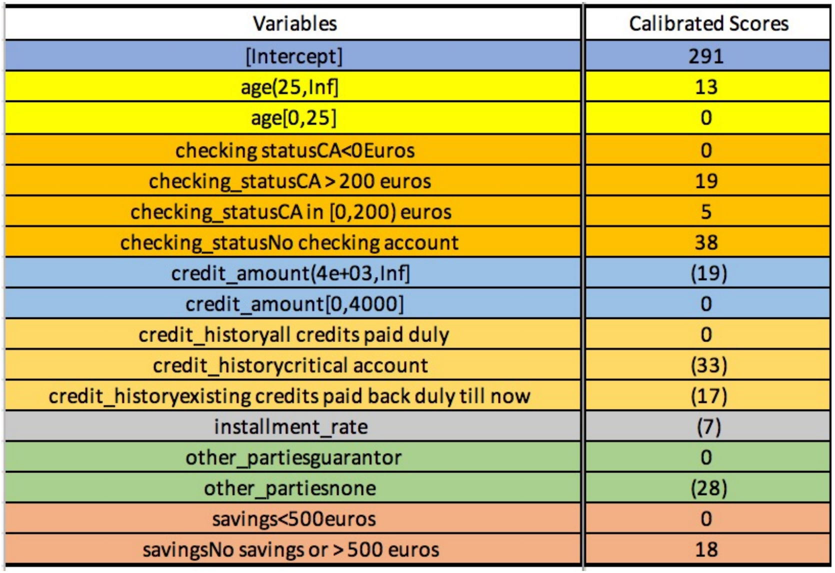
Binary Classification for Credit Scorecards

class = 1 if the account will default (bad account)

class = 0 if the account will not default (good account)



Sample Credit Scorecard



Variables	Calibrated Scores
[Intercept]	291
age(25,Inf]	13
age[0,25]	0
checking statusCA<0Euros	0
checking_statusCA > 200 euros	19
checking_statusCA in [0,200) euros	5
checking_statusNo checking account	38
credit_amount(4e+03,Inf]	(19)
credit_amount[0,4000]	0
credit_historyall credits paid duly	0
credit_historycritical account	(33)
credit_historyexisting credits paid back duly till now	(17)
installment_rate	(7)
other_partiesguarantor	0
other_partiesnone	(28)
savings<500euros	0
savingsNo savings or > 500 euros	18

Our model so far…

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 \mathrm{WOE}_{1,i} + \beta_2 \mathrm{WOE}_{2,i} + \cdots + \beta_k \mathrm{WOE}_{k,i}$$

bad : good odds

We want to apply a linear transformation to the log odds so that we'll have a scorecard of the form

$$\mathrm{Score}_i = \text{Base Score} + \sum_{m=1}^{k} \mathrm{Points}_{m,i}$$

The higher the credit score, the higher the probability of being creditworthy!

From probabilities to credit scores

$$\text{Score} = \text{Offset} + \text{Factor} \times \ln\left(\underbrace{\frac{1-\pi_i}{\pi_i}}_{\text{good : bad odds}}\right)$$

Take note that $\ln\left(\frac{1-\pi_i}{\pi_i}\right) = -\ln\left(\frac{\pi_i}{1-\pi_i}\right) = -\boldsymbol{X}^T\boldsymbol{\beta}$.

How do we find the offset and factor values?

The scorecard team (or the management) specifies the following:

1. the desired good : bad odds at a certain score s (odds_s, s)

2. the additional points p_2 required to double the good : bad odds $(2 \times \text{odds}_s, s + p_2)$

Solving for the offset and factor values

$$\text{Score} = \text{Offset} + \text{Factor} \times \ln(\text{odds})$$

Given the two given points (odds_s, s) and $(2 \times \text{odds}_s, s + p_2)$, it can be shown that

$$\text{Factor} = \frac{p_2}{\ln 2}$$

$$\text{Offset} = s - \text{Factor} \times \ln(\text{odds}_s)$$

Score scaling equation

$$
\begin{aligned}
\text{Score}_i &= \text{Offset} + \text{Factor} \times \ln(\text{odds}_i) \\
&= \text{Offset} + \text{Factor} \times \ln\left(\frac{1-\pi_i}{\pi_i}\right) \\
&= \text{Offset} + \text{Factor} \times \left[-\ln\left(\frac{\pi_i}{1-\pi_i}\right)\right] \\
&= \text{Offset} - \text{Factor} \times \left[\beta_0 + \beta_1 \text{WOE}_{1,i} + \beta_2 \text{WOE}_{2,i} + \cdots + \beta_k \text{WOE}_{k,i}\right] \\
&= [\text{Offset} - \text{Factor} \times \beta_0] + \sum_{m=1}^{k} -\text{Factor} \times \beta_m \times \text{WOE}_{m,i}
\end{aligned}
$$

Base Score $\qquad$ $\text{Points}_{m,i}$

Finding the cut-off score

Previously, we were able to find the optimal threshold z^* using the ROC curve. We use this threshold to determine if an account is good or bad.

$$\text{predicted class} = \begin{cases} 1, & \pi \geq z^* \\ 0, & \pi < z^* \end{cases}$$

We can use the score scaling equation to determine the **cut-off score** s^* for our scorecard. That is,

$$\text{predicted class} = \begin{cases} 1, & \text{score} \leq s^* \\ 0, & \text{score} > s^* \end{cases}$$

The cut-off score s^* is just the score at $\pi = z^*$.

Variable Selection

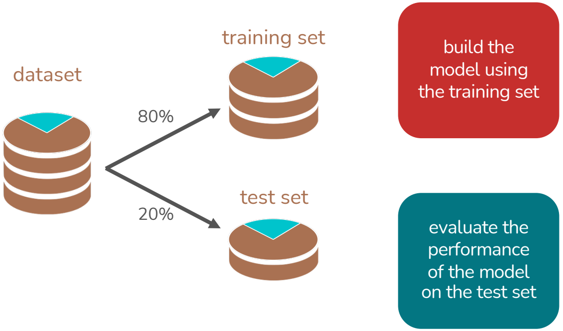
MATH 271: Statistical Methods

First Semester, S.Y. 2021-2022

Ateneo de Manila University

This session is recorded.

Modeling Workflow



dataset

80%

20%

training set

test set

build the model using the training set

evaluate the performance of the model on the test set

How can I improve the performance of my model?

What you can do for your Credit Scorecard Project:

1. Change your WOE bins.
2. Choose a good subset of the variables.

What you can do in the future, if you have the leeway:

1. Add more data! (more observations and/or more variables)
2. Check your imputation for missing values.
3. Use other data preprocessing procedures aside from WOE binning.
4. Change the model.
5. If the model is overfitting, apply some regularization techniques.

Variable Selection Procedures

Not all variables are useful in building a logistic regression model. You can select a subset of the variables and use this subset as inputs to your model.

If the dataset has *k* variables, how many subsets can you form?

You can form 2^k subsets! Thus, you can have 2^k different models! We need procedures to systematize the search for the best model from this huge space!

1. Significance of the variables ✓
2. Akaike Information Criterion (AIC)
3. Information Value

You could also test for multicollinearity in credit scorecards!

Variable Selection Procedures

1. Significance of the variables. For each parameter estimate $\widehat{\beta}_j$, we want to test whether the estimate is statistically significant. We have the following test.

$$H_0 : \beta_j \underline{= 0} \quad \text{vs} \quad H_a : \beta_{\underline{j \neq 0}}$$

If the $p -$ value given is less than a given level of significance, α, then we reject the null hypothesis and therefore the parameter estimate is significant.

Variable Selection Procedures

2. Akaike Information Criterion: The Akaike Information Criterion is a penalized log-likelihood criterion, which means that it measures goodness-of-fit but penalizes the complexity of the model. The AIC is given by

$$\text{AIC} = -2l + 2k$$

Where l is the log-likelihood function $\ln(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}(Y_i \ln(\hat{\pi}_i) + (1 - Y_i)\ln(1 - \hat{\pi}_i))$, and k is the number of parameters.

A better model is the one with the lower AIC.

Variable Selection Procedures

3. Information Value: Even before obtaining the parameters, some variables can be eliminated by taking note of the information value.

Selection of Variables

The selection of variables performed in R mainly makes use of variable significance or the AIC. Here are three methods:

1. **Forward Selection**
2. **Backward Selection**
3. **Stepwise Selection**

While these are statistical methods to obtain the best model, ultimately, a big influence on the selection of variables will be on expert judgment. Some variables are deemed by senior management as crucial to the decision of credit-worthiness.

Forward Selection

Forward Selection: The process starts with the null model, which is the model consisting of only an intercept term. Then the software determines which of the set of independent variables is the most statistically significant, given a significance level α. This variable is then added to the model, and the model parameters are refitted. The process continues until there are no more variables that are statistically significant.

On the basis of AIC, the model adds the variable such that when the model is refitted, the smallest AIC is obtained. This process is repeated until the it is better to stop adding variables than to add more, in terms of the AIC (i.e. if the current model has the lowest AIC compared to when one more variable is added to the model).

Backward and Stepwise Selection

Backward Selection: In this procedure, all variables are initially included in the model. Then one-by-one, the software removes the variables that are least statistically significant, or if the removal results to a lower AIC. This process is repeated until all variables are statistically significant, or if the minimal AIC has been attained.

Stepwise Selection: Stepwise selection incorporates both forward and backward selection processes by removing or adding variables to the model until the “best” model has been constructed. The process ends when the incremental predictive power of adding the next variable is negligible, or if the optimal AIC has already been achieved with the current model.

Multicollinearity in Credit Scorecards

- A lot of independent variables in credit scorecards.
- Issue: we might include redundant/include variables i.e, some variables can be written as a linear combination of others.
- The individual effect of a single variable (if it also depends on other variables in the model) cannot be properly extracted.

Variance Inflation Factor (VIF)

Problem with multicollinearity: parameter estimates will have large variances and covariances.

It describes the speed at which variances and covariances are increasing. The VIF is obtained by regressing an independent variable with the other independent variables and taking its adjusted R-square R^2, the VIF for each predictor is computed as

$$\mathrm{VIF}_j = \frac{1}{1 - \widehat{R_k}^2}$$

If $\mathrm{VIF}_j > 5.0$, this indicates a severe multicollinearity problem. These variables should be excluded from the model. If there is a group of independent variables with high VIFs, then include only one of those variables in the final model.

Other Model Validation Measures

MATH 271: Statistical Methods

First Semester, S.Y. 2021-2022

Ateneo de Manila University

This session is recorded.

Goodness of Fit: **Hosmer-Lemeshow Test**

We test the hypotheses:

H_0: The model is a good fit vs H_a: The model is not a good fit

The observations are first ranked in increasing order of $\widehat{\pi}_i$. Then two possible schemes can be carried out to categorize the data into deciles

Goodness of Fit: **Hosmer-Lemeshow Test**

- The first group will contain the observations in the bottom 10%, the second group will contain the observations in the second 10%, and so on. Cut offs are determined on the probabilities; that is, observations with an estimated probability of less than the first cut off will fall into the first group, and so on.

Goodness of Fit: **Hosmer-Lemeshow Test**

		Probability of Default
1	1	1.48%
	2	7.49%
2	3	21.38%
	4	24.97%
3	5	31.09%
	6	56.52%
4	7	56.57%
	8	60.00%
	9	63.05%
	10	64.36%
	11	65.33%
	12	66.81%
	13	68.98%
	14	71.51%
	15	86.79%
	16	87.07%
	17	89.85%
	18	91.41%
	19	93.55%
	20	95.31%

$O_4 = 1$

$N_4 = 2$

$\bar{\pi}_4 = 56.57\%$

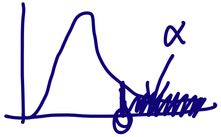
The H-L statistic can be computed through

$$\chi^2_{HL} = \sum_{j\ \boxed{i}=1}^{10} \frac{\left(O_j - N_j\bar{\pi}_j\right)^2}{N_j\bar{\pi}_j\left(1-\bar{\pi}_j\right)}$$

where N_j is the number of observations in the j^{th} decile, O_j is the number of bad accounts in the j^{th} decile, $\bar{\pi}_j$ is the average probability among bad accounts in the j^{th} decile.

Here, $\chi^2_{HL} \sim \chi(\overset{\downarrow}{8})$. Hence, the null hypothesis is rejected if $\chi^2_{HL} > \chi_\alpha(8)$ or if the p-value is less than α

$\alpha = 0.05$



Association Measures: Somer’s D

Generally, if account i is “good” while account j is “bad”, then we should expect that $\widehat{\pi}_i < \widehat{\pi}_j$. A good model must associate higher default probabilities to bad accounts than good accounts.

To measure this, we use **Somer’s D**

Association Measures: Somer's D

Consider all the possible $T = n_G \times n_B$ pairings of a good and bad account and the computed $\widehat{\pi}_i$ for each account.

Classify each pair as

- **Concordant** if $\widehat{\pi_B} > \widehat{\pi_G}$ (they agree on the rating)
- **Discordant** if $\widehat{\pi_B} < \widehat{\pi_G}$ (they disagree)
- **Tied** if $\widehat{\pi_B} = \widehat{\pi_G}$

The **Somer's D** statistic is given by

$$SD = \frac{(n_C - n_D)}{T}$$

Where n_C and n_D are the number of concordant and discordant pairs respectively.

Association Measures: Somer's D

11 bad accounts

9 good accounts

concordat

discordat

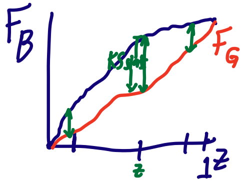
	Probability of Default
1	1.48%
2	7.49%
3	21.38%
4	24.97%
5	31.09%
6	56.52%
7	56.57%
8	60.00%
9	63.05%
10	64.36%
11	65.33%
12	66.81%
13	68.98%
14	71.51%
15	86.79%
16	87.07%
17	89.85%
18	91.41%
19	93.55%
20	95.31%

Exercise: Give an example of a concordant pair and discordant pair from the given:

Empirical Distribution and Cut-off Probabilities

Let B and G be the set of bad accounts and good accounts respectively, with n_B and n_G observations each.

The **empirical cumulative distribution of bad accounts** is given by



$$F_B(z) = \frac{1}{n_B} \sum_{i \in B} \mathbf{1}(\hat{\pi}_i \leq z)$$

This describes the proportion of bad accounts that are predicted to be good if the cut off probability is $z \in [0,1]$.

The **empirical cumulative distribution of good accounts** is given by

$$F_G(z) = \frac{1}{n_G} \sum_{i \in G} \mathbf{1}(\hat{\pi}_i \leq z)$$

Which describes the proportion of good accounts that are predicted to be good if the cut-off probability is $z \in [0,1]$.

Kolmogorov-Smirnov (KS) statistic

We want to choose the cut-off z such that the empirical cdfs of the good and bad accounts are as different from each other as possible. That is, our cut-off probability is the value z that maximizes $|F_B(z) - F_G(z)|$.

The maximum value, given by $KS = \sup_z \boxed{|F_B(z) - F_G(z)|}$ is actually the **Kolmogorov-Smirnov (KS) statistic,** which is used to determine whether the distribution of the "bad" accounts is significantly different from the distribution "good" accounts.

Kolmogorov-Smirnov (KS) statistic

We test:

H_0: The two distirbutions are the same vs H_a: The two distributions are not the same

We reject H_0 at the significance level of α if

$$KS \geq k(\alpha)\sqrt{\frac{1}{n_B}+\frac{1}{n_G}}$$

With $k(\alpha) = \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}$

Cumulative Accuracy Profile (CAP)

The CAP visually describes the descriptive power of a model. The y-axis is the 1-ECDF of the bad accounts and the x-axis is the 1-ECDF of all accounts, i.e the CAP is the graph of the points $\left(1 - F(z), 1 - F_B(z)\right)$ as z ranges from [0,1]

The empirical distribution of all accounts is given by

$$F(z) = \frac{1}{n_B + n_G} \sum_{i \in B \cup G} \mathbf{1}(\hat{\pi}_i \leq z)$$

Cumulative Accuracy Profile (CAP)

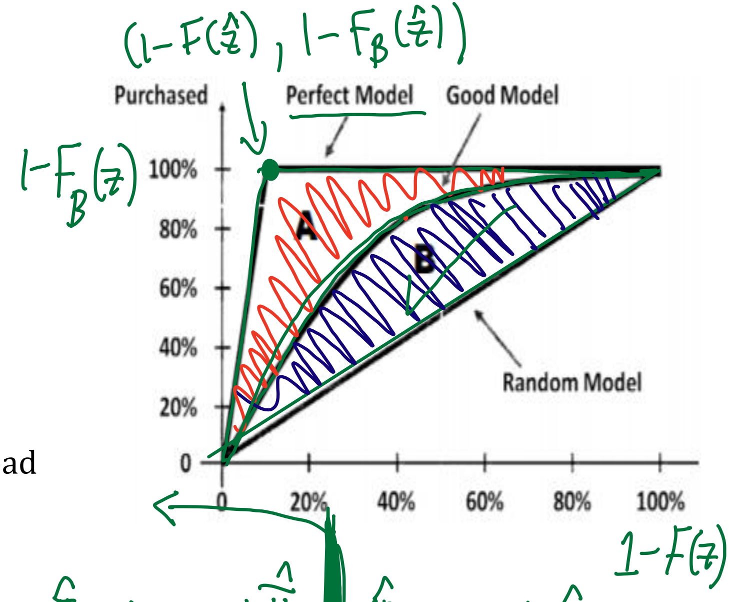
We are plotting

$$\underline{1 - F(z)} = \frac{1}{n_B + n_G} \sum_{i \in B \cup G} \mathbf{1}(\hat{\pi}_i > z) \quad \text{vs}$$

$$\underline{1 - F_B(z)} = \frac{1}{n_B} \sum_{i \in B} \mathbf{1}(\hat{\pi}_i > z)$$

A perfect model is such that there is a cut off probability that splits all accounts with the bad accounts. The accuracy ratio is given by the formula

$$\boxed{AR} = \frac{B}{A + B}$$



$$\hat{\pi}_1 < \hat{\pi}_2 < \dots < \hat{\pi}_k < \hat{\pi}_{k+1} \dots < \hat{\pi}_n$$

$\hat{z}$

Relationships Between CAP, Gini, and the AUC

In general, the accuracy ratio obtained from the CAP is just equivalent to the Gini coefficient. Furthermore,

$$\boxed{\underline{AUC} = \frac{1 + Gini}{2}} \checkmark$$

① HL stat

② Somer's D

③ KS stat

④ AR fm CAP = Gini coefficient