Credit Risk

panko.aliaksandr

October 2020

Contents

1 Introduction

Credit risk is a risk of a counter-party default.

First read this great [article]

2 Judgmental Method: 5 Cs of Credit

The "5 Cs of Credit" is a common phrase used to describe the five major factors used to determine a potential borrower's creditworthiness. The 5 Cs of Credit refer to Character, Capacity, Collateral, Capital, and Conditions.

3 Core Definitions

1. PD - probability of default

2. EAD - exposure at default (the amuont of the remaining debt on the default date)

3. LGD - loss given default = 1-recovery rate

4. EL - expected loss. EL = PD * EAD * LGD
5. M - Effective Maturity

4 Modeling

1. Define factors
2. Prepare data
3. Split data into Train and Test
4. Create a model
5. Analyse a confusion matrix (matrix with true positives and negatives and false positives and negatives, basically type 1 and 2 errors)

5 Model Fit

5.1 Part 1: Sample probability

- Let us define a random variable Y, which takes 2 values: default (1) and not default(0).
- Since random variable takes only 2 values $Y \sim Bernoulli(p)$
- Assume in the sample there are n observations. In this case we say that we have n different random variables $Y_1, Y_2, ..., Y_n$.
- Different since it is clear that based on factors values $X = (x_1, x_2...x_n)$ the parameter of Bernoulli distribution is different in each case, meaning that $p_1 = p_1(x_1, x_2...x_n), p_2 = p_2(x_1, x_2...x_n), etc.$
- This means we have

  $$Y_1 \sim B(p_1), Y_2 \sim B(p_2), ..., Y_n \sim B(p_n)$$

  or $Y \sim B_{joint}(p_1, p_2, .., p_n)$
- $PDF_{Bernoulli} = p^k(1-p)^{1-k}, k \in (0,1)$
- The probability of observing this sample is:

  $$P(sample) = \prod_i p_i^{k_i}(X, \Theta)(1 - p_i^{1-k_i}(X, \Theta))$$

- We know $X = (x_1, x_2...x_n)$ and k_i for each run, but we do not know the dependency (formula) between $X = (x_1, x_2...x_n)$ and $p(X)$. If we construct such formula $p_i(X, \Theta)$, we can use MLE approach to find parameters $\Theta = (\beta_0, \beta_1, ...\beta_m)$

5.2 Transformation

So the task is to construct a model which could map factors $X = (x_1, x_2...x_n)$ and probability of default $p(X)$.

- First step is to make an important assumption about the model type. Assuming that the factors influence probability of default in linear way, we can construct the formula using linear regression.

- However, 'standard/normal' linear regression analysis required dependent variable to be unrestricted numerical $y \in [-Inf; Inf]$. In our case $p(x) \in [0; 1]$. This means that we cannot use OLS directly, transformation is required.

- Notice that $\frac{P}{1-P}$ belongs to [0; Inf].

- $\ln(\frac{p}{1-p})$ belongs to $[-Inf; Inf]$. This term is called log-odds.

- Now we can operate with log-odds dependent variable as usually and fit a regression line:

$$\ln(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$$

- Making the assumption that factors influence the log-odds in a linear way we can conclude that (applying exponent operation and finding p from the ratio): $p(x) = \frac{1}{1+e^{b_0+b_1x_1+...b_nx_n}} = \frac{1}{1+e^{\sum_{i=0:n}\beta_i x_i}}$, having in mind $x_0 = 1$ in this view.

- Now we are ready to combine 2 parts together:

$$P(sample) = \prod_i \frac{1}{1+e^{\sum_i \beta_i x_i}}^{k_i} (1 - \frac{1}{1+e^{\sum_i \beta_i x_i}}^{1-k_i})$$

- Now we have all k_i (default(1) or not(0)), all x_i (factors values), and we can apply MLE approach and maximize log of this function to find parameters β_i. After that the model is fitted. Using the regression we can find the probability of default. Having the probability we can make a decision.