Ordinal Methods for Corporate Credit Rating Classification

Rodrigo Palmaka

The Data

Features

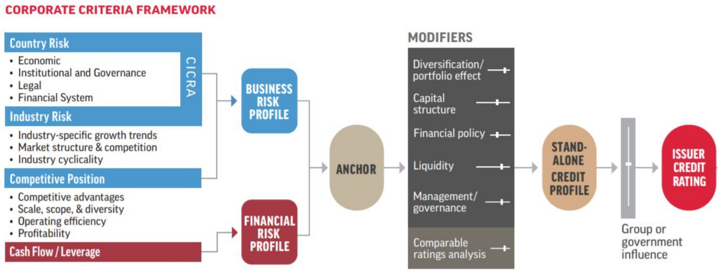
	Variables
X1	Enterprise Value Multiple
X2	P/E (Diluted, Excl. EI)
X3	Price/Cash flow
X4	Net Profit Margin
X5	Operating Profit Margin Before Depreciation
X6	Cash Flow Margin
X7	Total Debt/Invested Capital
X8	Cash Balance/Total Liabilities
X9	Total Debt/EBITDA
X10	Profit Before Depreciation/Current Liabilities
X11	Operating CF/Current Liabilities
X12	Cash Flow/Total Debt
X13	Total Liabilities/Total Tangible Assets
X14	Total Debt/Capital
X15	Total Debt/Equity
X16	Cash Ratio
X17	Quick Ratio (Acid Test)
X18	Price/Book
X19	Average Sales Price - NGL
X20	Average Sales Price - NG
X21	Average Sales Price - Oil
X22	Production - NGL (Total)
X23	Production - NG (Total)
X24	Production - Oil (Total)
X25	Dry Hole Expense
X26	Exploration Expense

Labels

Credit Rating Scales by Agency, Long-Term

Moody's	S&P	Fitch	
Aaa	AAA	AAA	Prime
Aa1	AA+	AA+	High grade
Aa2	AA	AA	
Aa3	AA-	AA-	
A1	A+	A+	Upper medium grade
A2	A	A	
A3	A-	A-	
Baa1	BBB+	BBB+	Lower medium grade
Baa2	BBB	BBB	
Baa3	BBB-	BBB-	
Ba1	BB+	BB+	Non-investment grade speculative
Ba2	BB	BB	
Ba3	BB-	BB-	
B1	B+	B+	Highly speculative
B2	B	B	
B3	B-	B-	
Caa1	CCC+	CCC	Substantial risk
Caa2	CCC		Extremely speculative
Caa3	CCC-		Default imminent with little prospect for recovery
Ca	CC	CC	
	C	C	
C	D	D	In default
/			
/			

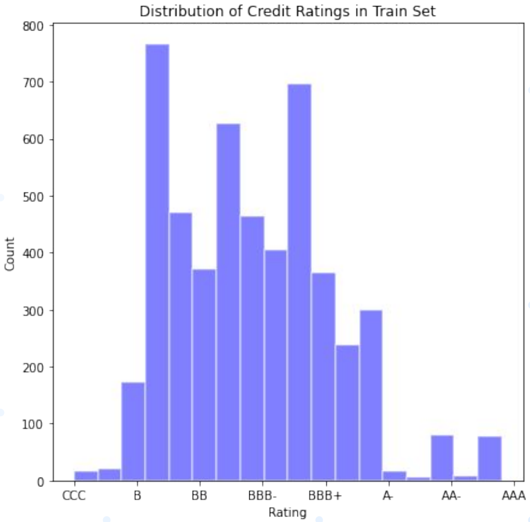
"Junk"



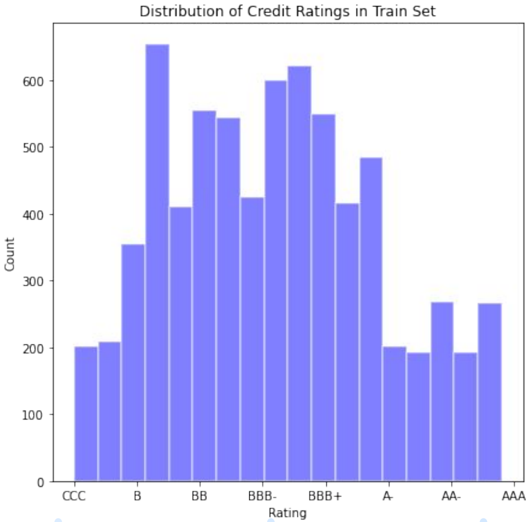
S&P corporate rating pipeline

- Data from 97 Energy companies
- 2006 - 2017

Label Distribution



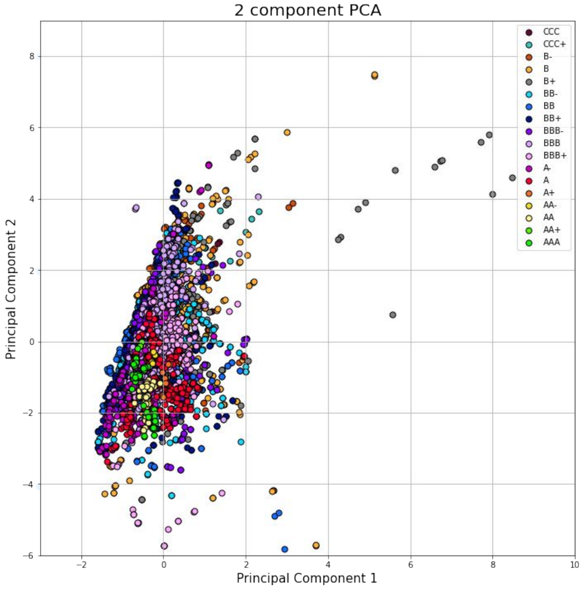
Distribution of Labels for the Train Set



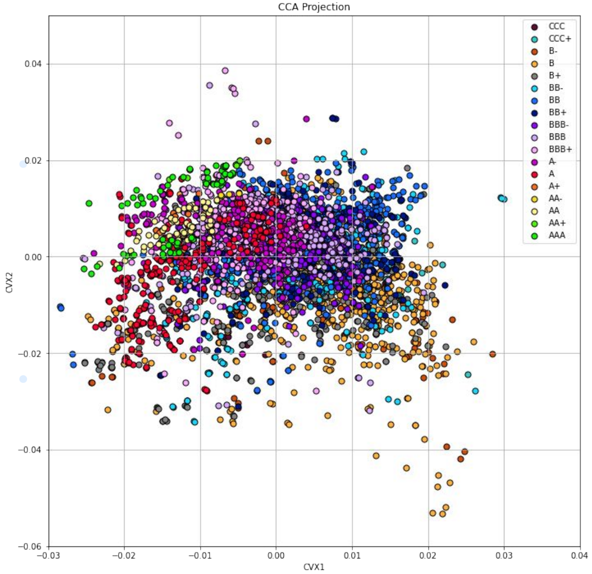
Distribution of SMOTE Labels for the Train Set

Chawla et al., 2002. *SMOTE: Synthetic Minority Over-sampling Technique*

Dimensionality Reduction



PC1, PC2 for full label set



CCA for full label set

Baseline Models

LDA/QDA	Random Forest	SVM	FC Neural Network	LSTM
Regularization: 1×10^{-4}	Max depth: 11-13 Max features: 10	C: 1000-2000 Kernel: RBF Gamma = 0.1	lr: 7×10^{-4} Hidden layers: 5 Hidden size: 250	lr: 1×10^{-3} Hidden size: 250 Sequence len.: 6

Ordinal Regression Motivation

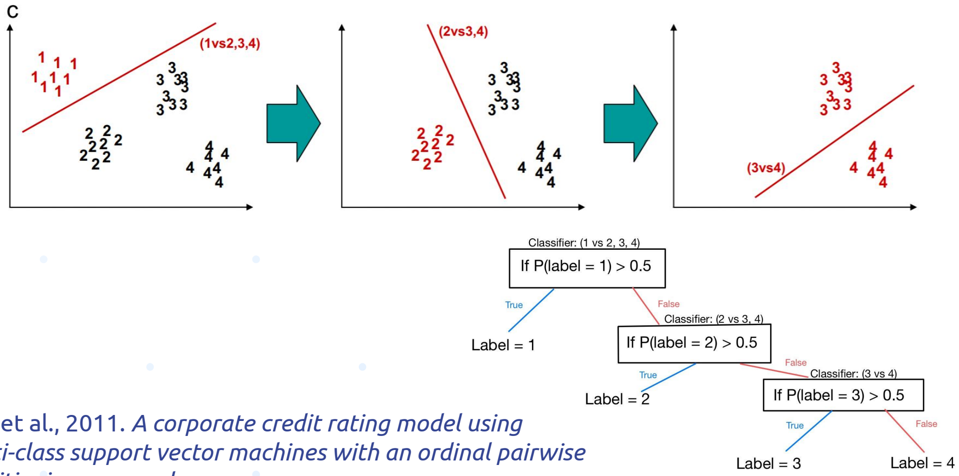
- Classifiers such as SVM don’t consider intrinsic ordering of target variables.
- Example: determining if a data point is green, blue or red vs. deciding if data is cold, warm or hot.
- Naturally, for credit ratings we have a total order:
  AAA > AA+ > … > BB+ > … > CCC+ > CCC

Ordinal Models

Ordinal SVM

OMSVM

- C - 1 binary classifiers
- "One-Against-Followers", forward direction



Ahn et al., 2011. *A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach*

SORSVM

- C - 1 binary classifiers
- Each learns probability of data point being greater than rating {r_1, …, r_C-1}

$$P(r_1) = 1 - P(label > r_1)$$

$$P(r_k) = P(label > r_{k-1}) - P(label > r_k), \quad 1 < k < C$$

$$P(r_C) = P(label > r_{C-1})$$

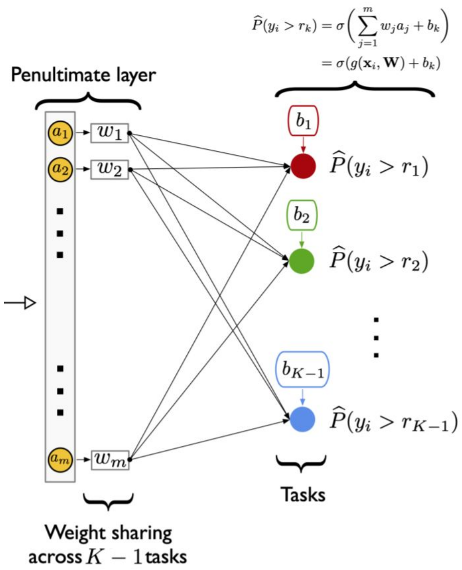
We assign the label with highest probability to the test point:

$$\hat{f}(\boldsymbol{x_{test}}) = \arg\max_k P(r_k)$$

Frank et al., 2001. *A Simple Approach to Ordinal Classification*

CORAL NN

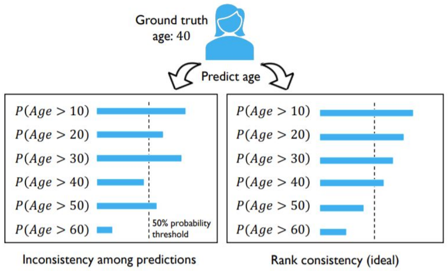
- Ordinal One-Hot Encoding
- Guarantees rank consistency
- Sigmoid output layer
- Shared weights, independent biases in penultimate layer
- Custom Cross Entropy loss



label classification problem: We obtain the multi-label target vector $\mathbf{y} = [y_1, \ldots, y_K] \in \{0,1\}^K$ from r such that

$$y_j = \begin{cases} 0 & j < k \\ 1 & j \geq k \end{cases} \quad (2)$$

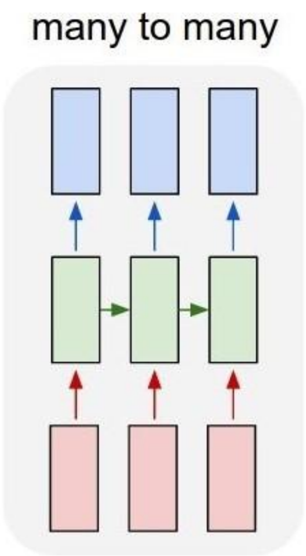
where $j = 1, 2, \ldots, K$.

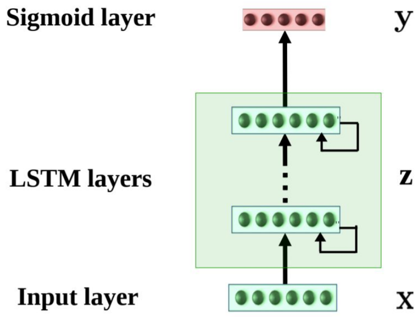


Cao et al., 2020. *Rank consistent ordinal regression for neural networks with application to age estimation*

LSTM-OR

- Ordinal One-Hot Encoding
- Many-to-many LSTM
- Sigmoid output layer
- Allows us to consider information (hidden state) from previous time step





Vishnu TV et al., 2019. *Data-driven Prognostics with Predictive Uncertainty Estimation using Ensemble of Deep Ordinal Regression Models*

Evaluation Methodology

Random Sampling

- Sample data points randomly independent of time period and company.
- The prevailing method used in corporate credit ratings research.
- Use data from the future to predict the past.
- Samples aren't i.i.d. and actually heavily dependent (ratings don't change too often)
- Learns generalized representation of financial data to rating independent of macro trends.

Time Series Splits (LSTM models only)

- Test set comprised of data from periods not seen in training data.
  - Last 6 month window for each company used for testing.
- Using previous data points recurrently can cause model to simply output the previous rating.
- No mixing of past and present data.
- Learns to model ratings changes over time.

Results - Random Sampling

	Accuracy	MAE
LDA	30.99	----
QDA	41.14	----
Random Forest	**95.36**	**0.09**
SVM	**90.27**	**0.22**
SOR-SVM	87.77	0.25
OMSVM	86.45	0.33

	Accuracy	MAE
Neural Network	**88.58**	0.237
CORAL NN	85.70	**0.20**
LSTM	**93.94**	0.13
LSTM-OR	92.53	**0.12**

Note: LSTM models use 6-period window

Results - Time Series

	Accuracy	MAE
LSTM	**58.90**	1.18
LSTM-OR	50.60	**1.16**

Conclusions

- SMOTE unable to improve accuracy
- Some ordinal models reduce MAE
- Need to consider separating companies in train/test sets
- More information can sometimes mean more constraints

Contact

rodrigopalmaka@gmail.com

github.com/rodrigo-palmaka/honors-thesis

linkedin.com/in/rodrigopalmaka