Predicting Long-Term Unemployment

A Comparison Between Logistic Regression and Kernel SVM

Malik Stromberg

29/01/2021

University of Tübingen

Outline

Introduction

Consequences of Unemployment

- A job-loss affects
  - Psychological well-being [15]
  - Perceived life quality [15]
  - Career quality [12]
- The lacks increase in unemployment duration
- Reemployment helps to mitigate this decline [1]

$\rightarrow$ **It is desirable to find a new job quickly.**

But some unemployed people need specific state aid to get reemployed.

Can those individuals be identified more accurately using machine learning methods in comparison to traditional approaches?

Overview

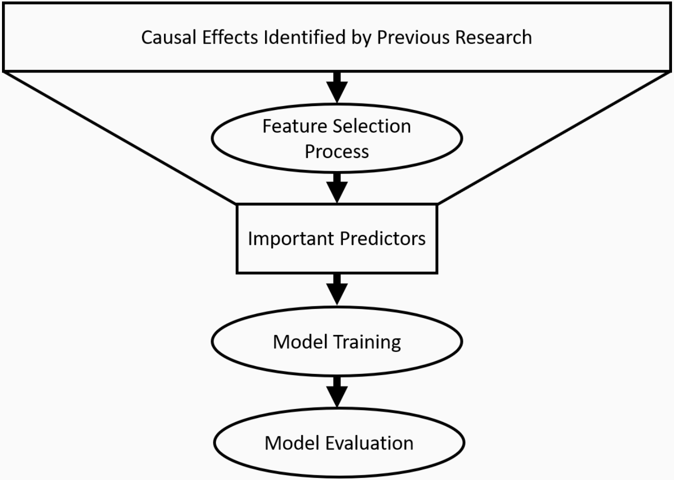

Figure 1: Workflow overview.

Context

Long-Term Unemployment as a Multidisciplinary Field

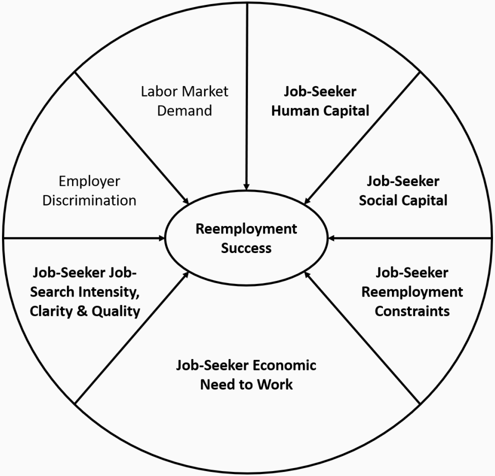

Figure 2: Conceptual model of variable groups associated with reemployment success.

Wanberg, Hough, and Song [20]

Causal Effects on Long-Term Unemployment

Publication	Country	Method	Effects
Wanberg, Watt, and Rumsey [21]	USA	Logit	Age Job seeking support
Obben, Engelbrecht, and Thompson [17]	New Zealand	Logit	Gender Ethnicity Schooling Regional location
Marelli and Vakuenko [13]	Russia, Italy	Probit	Marital status Health Household income
O'Connell, McGuinness, and Kelly [16]	Ireland	Probit	Employment history Willingness to move Unemployment duration Public transports
Curtis, Gibbon, and Katsikitis [3]	Australia	ANOVA	In-group identification
Kokko, Pulkkinen, and Puustinen [8]	Finland	MCA	Depressive symptoms Self-control
Lötters et al. [11]	Netherlands	Logit	Perceived Health Willingness to accept a job
Lallukka et al. [10]	Finland	Logit	Social determinants
Krause [9]	Germany	Logit	Satisfaction
Wanberg, Hough, and Song [20]	USA	Logit	Children in household Economic hardship
De Battisti et al. [4]	Italy	SEM	Perceived employability Psychological distress

Table 1: Causal effects on long-term unemployment suggested by past literature.

JSCI-Score

Long-term unemployment prediction model
developed by Matty [14]

Goal:	Develop a score that gives information about the risk of an unemployed individual becoming long-term unemployed.
Country:	United Kingdom
Data:	Questionnaire
Feature Selection:	Trying different combinations of predictors
Variables:	Individual's economic need to work Reemployment constraints Attitudinal variables Administrative variables
Model:	Logistic regression
Classification:	JSCI-Score

Method

General ERM Framework

Input space:	$\mathcal{X}$
Actual output space:	$\mathcal{Y}_{act}$
Prediction output scpace:	$\mathcal{Y}_{pred}$
Training data points:	$(x_i, y_i)_{i=1,\ldots,n} \in \mathcal{X} \times \mathcal{Y}_{act}$
Function space:	$\mathcal{F} : \mathcal{X} \to \mathcal{Y}_{pred}$
Loss function:	$\ell : \mathcal{X} \times \mathcal{Y}_{act} \times \mathcal{Y}_{pred} \to \mathbb{R}_{\geq 0}$

Shalev-Schwartz and Ben-David [18]

General ERM Framework

The function with the smallest empirical risk is given by

$$f_n := \operatorname*{argmin}_{f \in \mathcal{F}} R_n(f) \tag{1}$$

with empirical risk

$$R_n(f) := \frac{1}{n} \sum_{i=1}^{n} \ell(x_i, y_i, f(x_i)). \tag{2}$$

Shalev-Schwartz and Ben-David [18]

Regularized ERM Framework

Regularizer: $\Omega : \mathcal{F} \to \mathbb{R}_{\geq 0}$
Hyperparameter: $\lambda \in \mathbb{R}$

The regularized empirical risk is given by

$$R_{reg,\,n}(f) := R_n(f) + \lambda\,\Omega(f). \tag{3}$$

Shalev-Schwartz and Ben-David [18]

Methodological Procedure

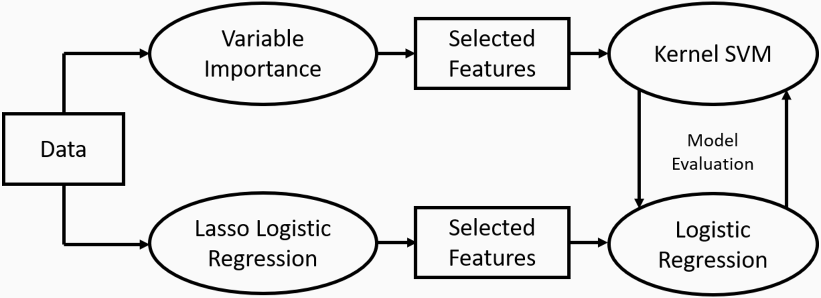

Figure 3: Methodological procedure.

Logistic Regression with ERM

Number of selected features:	d
Input space:	$\mathcal{X} = \mathbb{R}^d$
Actual output space:	$\mathcal{Y}_{act} = \{\pm 1\}$
Prediction output space:	$\mathcal{Y}_{pred} = \mathbb{R}$
Function space:	$\mathcal{F} = \{f(x_i) = \langle w, x_i \rangle;\ w \in \mathbb{R}^d\}$

Define the logistic loss as

$$\ell_{log}(x_i, y_i, f(x_i)) := log(1 + exp(-y_i f(x_i))). \tag{4}$$

Shalev-Schwartz and Ben-David [18]

Logistic Regression with ERM

The logistic ERM problem is given by

$$\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} log\left(1 + exp\left(-y_i \langle w, x_i \rangle\right)\right). \tag{5}$$

A new data point $x_{new} \in \mathbb{R}^d$ can be classified using

$$P(x_{new}) = \frac{1}{1 + exp(-\langle w, x_{new} \rangle)}, \tag{6}$$

$$y^{label} = \begin{cases} 1 & \text{if} \quad P(x_{new}) \geq 0.5, \\ -1 & \text{if} \quad P(x_{new}) < 0.5. \end{cases} \tag{7}$$

Shalev-Schwartz and Ben-David [18]

Feature Selection with Lasso Logistic Regression

Number of considered features: D

Defining the lasso regularizer

$$\Omega_{lasso}(f) := ||w||_1 \tag{8}$$

leads to the regularized logistic ERM problem

$$\underset{w \in \mathbb{R}^D}{\operatorname{argmin}} \frac{1}{m} \sum_{i=1}^{m} log\left(1 + exp\left(-y_i \langle w, x_i \rangle\right)\right) + \lambda \, ||w||_1. \tag{9}$$

Tibshirani [19]

SVM with ERM

Number of selected features:	d
Input space:	$\mathcal{X} = \mathbb{R}^d$
Actual output space:	$\mathcal{Y}_{act} = \{\pm 1\}$
Prediction output space:	$\mathcal{Y}_{pred} = \mathbb{R}$
Function space:	$\mathcal{F} = \{f(X) = \langle w, X \rangle + b;\ w \in \mathbb{R}^d, b \in \mathbb{R}\}$
Hyperparameter:	C

Define the hinge loss as

$$\ell_{hinge}(x_i, y_i, f(x_i)) := max\,\{0; 1 - y_i\, f(x_i)\} \tag{10}$$

and regularizer

$$\Omega_{ridge}(f) := ||w||^2. \tag{11}$$

Hastie, Tibshirani, and Wainwright [7]

SVM with ERM

The ERM primal problem for SVM is given by

$$\underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \; \frac{C}{n} \sum_{i=1}^{n} max \left\{0; 1 - y_i \, f(x_i)\right\} + ||w||^2. \tag{12}$$

Converting the problem to the corresponding dual problem yields

$$\max_{\alpha \in \mathbb{R}^n} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \, \alpha_j \, y_i \, y_j \, \langle x_i, x_j \rangle \tag{13a}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq \frac{C}{n} \tag{13b}$$

$$\sum_{i=1}^{n} \alpha_i \, y_i = 0. \tag{13c}$$

Hastie, Tibshirani, and Wainwright [7]

Kernel SVM

Kernel function: $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
Hyperparameter: σ

Define the Gaussian kernel as

$$k_{gauss}(x_i, x_j) := exp\left(- \frac{||x_i - x_j||^2}{2\,\sigma} \right) \tag{14}$$

leads to the optimization problem

$$\max_{\alpha \in \mathbb{R}^n} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i\, \alpha_j\, y_i\, y_j\, exp\left(- \frac{||x_i - x_j||^2}{2\,\sigma} \right) \tag{15a}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq \frac{C}{n} \tag{15b}$$

$$\sum_{i=1}^{n} \alpha_i\, y_i = 0, \tag{15c}$$

Cortes and Vapnik [2]

Kernel SVM

$$\textbf{Subspace}: \quad J = \{j \mid 0 < \alpha_j < \tfrac{C}{n}\}$$

The function value for new data point $x_{new} \in \mathbb{R}^d$ can be calculated as

$$f(x_{new}) = \sum_{i=1}^{n} \alpha_i \, y_i \, k(x_i, x_{new}) + \left(y_j - \sum_{i=1}^{n} y_i \, \alpha_i \, k(x_i, x_j) \right); \quad j \in J \quad (16)$$

and classified with

$$y^{label} = \begin{cases} 1 & \text{if} \quad f(x_{new}) \geq 0, \\ -1 & \text{if} \quad f(x_{new}) < 0. \end{cases} \quad (17)$$

Cortes and Vapnik [2]

Measuring Model Performance

1. Area under the ROC-curve
2. Emprical, test and CV risk with 0-1-loss:

$$\ell_{0-1}(x_i, y_i, y^{label}) := \begin{cases} 0 & \text{if} \quad y_i = y^{label}, \\ 1 & \text{if} \quad y_i \neq y^{label} \end{cases} \tag{18}$$

Guyon and Elisseeff [6]

Data

German Socio-Economic Panel

- Annual multidisciplinary survey
- 30,000 individuals
- Since 2009 it seamlessly records monthly unemployment data
- 2,524 labeled observations
- 30 variables considered, including
  - Demographic variables
  - Household income
  - Worries & satisfaction
  - Health
  - Social activities
  - Reemployment constraints & perceived employability
  - Character traits

diw.de [5]

Sample in Workflow

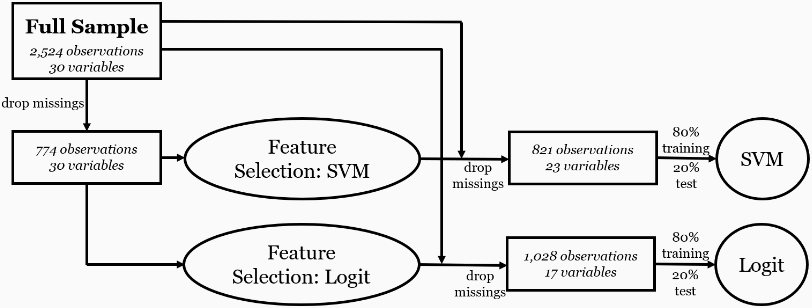

Figure 4: Methodological procedure with sample sizes and number of independent variables.

Results & Discussion

Lasso Logistic Regression

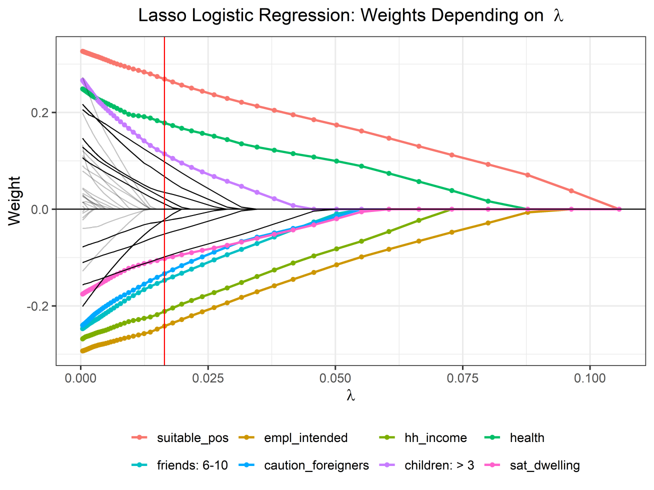

Figure 5: Predictor weights for logistic regression depending on regularization parameter λ.

SVM Variable Importance

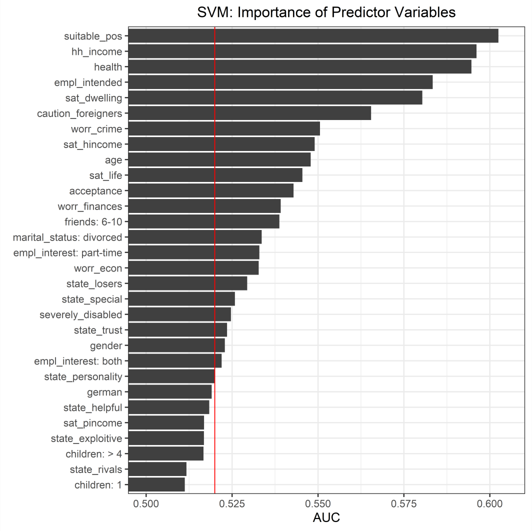

Figure 6: Predictor importance in SVM measured by AUC.

Feature Selection

<table>
<tr><th>Logistic Regression</th><th>SVM</th></tr>
<tr><td colspan="2">worr_crime</td></tr>
<tr><td colspan="2">sat_dwelling</td></tr>
<tr><td colspan="2">health</td></tr>
<tr><td colspan="2">caution_foreigners</td></tr>
<tr><td colspan="2">friends</td></tr>
<tr><td colspan="2">marital_status</td></tr>
<tr><td colspan="2">acceptance</td></tr>
<tr><td colspan="2">empl_intended</td></tr>
<tr><td colspan="2">state_personality</td></tr>
<tr><td colspan="2">hh_income</td></tr>
<tr><td colspan="2">suitable_pos</td></tr>
<tr><td colspan="2">empl_interest</td></tr>
<tr><td>german</td><td>worr_econ</td></tr>
<tr><td>children</td><td>worr_finances</td></tr>
<tr><td>state_helpful</td><td>sat_hincome</td></tr>
<tr><td></td><td>sat_life</td></tr>
<tr><td></td><td>gender</td></tr>
<tr><td></td><td>severely_disabled</td></tr>
<tr><td></td><td>age</td></tr>
<tr><td></td><td>state_losers</td></tr>
<tr><td></td><td>state_trust</td></tr>
<tr><td></td><td>state_special</td></tr>
</table>

Table 2: Selected features for logistic regression and SVM.

Performance

Measure	Logistic Regression	SVM
Empirical risk	0.3147	0.352
CV risk	0.3573	0.399
Test risk	0.3527	0.3879
Training AUC	0.74	0.7157
Test AUC	0.6628	0.732

Table 3: Performance measures for logistic regression and SVM.

ROC Curves

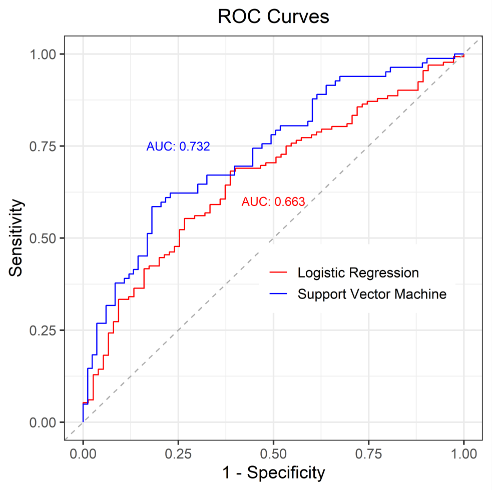

Figure 7: ROC curves on test data for logistic regression and kernel SVM.

Limitations

Limitations

- Remarkable decline in number of observations due to high missing rate
- Anonymized individual data
- In recent years critics arose about AUC as a single variable importance measure

Further Research

Further Research

- Other machine learning approaches
- Adjust loss-function resulting in a cost-sensitive classification
- Psychological consequences

Summary

Summary

Goal: Compare predictive power of logistic regression and kernel SVM for classification of unemployed individuals into classes of small and high risk of becoming long-term unemployed based on individual-specific features.

Method: Feature selection: lasso regularization, importance measures
Model: Logistic regression, kernel SVM

Data: German SOEP data, 2009-2017

Result: Kernel SVM can outperform traditional approaches in predictive power measured by AUC.

Thanks!

References

[1] Jennie E. Brand. “The far-reaching impact of job loss and unemployment”. In: *Annual Review of Sociology* 41 (2015), pp. 359–375.

[2] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine Learning* 20 (1995), pp. 273–297.

[3] Elli Curtis, Peter Gibbon, and Mary Katsikitis. “Group identity and readiness to change unemployment status”. In: *Journal of Employment Counseling* 53 (2016), pp. 50–59.

[4] Francesca De Battisti et al. "Perceived employability and reemployment: Do job search strategies and psychological distress matter?" In: *Journal of Occupational and Organizational Psychology* 89 (2016), pp. 813–833.

[5] diw.de. accessed 04-January-2021. URL: https://www.diw.de/en/diw_01.c.615551.en/research_infrastructure__socio-economic_panel__soep.html.

[6] Isabelle Guyon and André Elisseeff. "An introduction to variable and feature selection". In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182.

[7] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. *Statistical learning with sparsity: The lasso and generalizations*. CRC Press, Taylor & Francis Group, 2015. URL: https://web.stanford.edu/~hastie/StatLearnSparsity/index.html.

[8] Katja Kokko, Lea Pulkkinen, and Minna Puustinen. “Selection into long-term unemployment and its psychological consequences”. In: *International Journal of Behavioral Development* 24.3 (2000), pp. 310–320.

[9] Annabelle Krause. “Don’t worry, be happy? Happiness and reemployment”. In: *Journal of Economic Behavior & Organization* 96 (2013), pp. 1–20.

[10] Tea Lallukka et al. “Determinants of long-term unemployment in early adulthood: A Finnish birth cohort study”. In: *SSM - Population Health* 8 (2019), p. 100410.

[11] Freek Lötters et al. “The influence of perceived health on labour participation among long term unemployed”. In: *Journal of Occupational Rehabilitation* 23 (2013), pp. 300–308.

[12] Anna Manzoni and Irma Mooi-Reci. "The cumulative disadvantage of unemployment: Longitudinal evidence across gender and age at first unemployment in Germany". In: *PLoS ONE* 15.6 (2020), e0234786.

[13] Enrico Marelli and Elena Vakuenko. "Youth unemployment in Italy and Russia: Aggregate trends and individual determinants". In: *The Economic and Labour Realtions Review* 27.3 (2016), pp. 387–405.

[14] Simon Matty. "Predicting likelihood of long-term unemployment: The development of a UK jobseekers' classification instrument". In: *Working Paper No. 116, Department for Work and Pensions* (2013).

[15] Victoria Mousteri, Michael Daly, and Liam Delaney. "The scarring effect of unemployment on psychological well-being across Europe". In: *Social Science Research* 72 (2018), pp. 146–169.

[16] Philip J. O'Connell, Seamus McGuinness, and Elish Kelly. "The transition from short- to long-term unemployment: A statistical profiling model for Ireland". In: *The Economic and Social Review* 43.1 (2012), pp. 135–164.

[17] James Obben, Hans-Jürgen Engelbrecht, and Wesley Thompson. "A logit model of the incidence of long-term unemployment". In: *Applied Economics Letters* 9.1 (2002), pp. 43–46.

[18] Shai Shalev-Schwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2014. URL: http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning.

[19] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288.

[20] Connie R. Wanberg, Leatta M. Hough, and Zhaoli Song. "Predictive validity of a multidisciplinary model of reemployment success". In: *Journal of Applied Psychology* 87.6 (2002), pp. 1100–1120.

[21] Connie R. Wanberg, John D. Watt, and Deborah J. Rumsey. "Individuals without jobs: An empirical study of job-seeking behavior and reemployment". In: *Journal of Applied Psychology* 81.1 (1996), pp. 76–87.