Utility-based Regression

Luis Torgo, Paula Branco and Rita Ribeiro

Dalhousie Univ., Canada
INESC Tec, Univ. Porto, Portugal

May, 2018

DALHOUSIE UNIVERSITY

Predictive Analytics - 1

Standard Learning Tasks

- Goal: obtain a good approximation of an unknown function $Y = f(X_1, X_2, \cdots, X_p)$
- Maps a set of p predictors into a target variable Y, which may be numeric (**regression**) or nominal (classification)
- Use a training set $D = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{n}$ to obtain the approximation, i.e. the model, $h(X_1, X_2, \cdots, X_p)$

DALHOUSIE UNIVERSITY

Predictive Analytics - 1

Standard Learning Tasks

- Goal: obtain a good approximation of an unknown function $Y = f(X_1, X_2, \cdots, X_p)$
- Maps a set of p predictors into a target variable Y, which may be numeric (**regression**) or nominal (classification)
- Use a training set $D = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{n}$ to obtain the approximation, i.e. the model, $h(X_1, X_2, \cdots, X_p)$
- Model is obtained using some preference criterion
- On standard learning tasks this is usually some form of minimization of the **average prediction error**

Predictive Analytics - 2

Utility-based Learning Tasks

- End-user has a non-uniform preference over the target variable domain

DALHOUSIE UNIVERSITY

Predictive Analytics - 2

Utility-based Learning Tasks

- End-user has a non-uniform preference over the target variable domain
- Some errors may be more costly than others
- Some accurate predictions may bring more rewards than others

Challenges of Utility-based Learning

1. How to express user preferences?
2. Standard evaluation metrics may be misleading
3. How to bias learning algorithms to the cases the end user regards as more relevant?

DALHOUSIE UNIVERSITY

UBL for Classification

This is a well studied problem for classification tasks (nominal target)

- Adequate metrics have been proposed (e.g. precision/recall, etc.)
- Learning methods or resampling strategies were developed to optimize these metrics

UBL for Regression

Problem not so thoroughly studied for regression (numeric target)

Many relevant applications

- Forecasting extreme values of some ecological parameters
- Forecasting stock market returns
- etc.

An Example

Forecasting Stock Market Returns

- Very high or low returns (% variations of prices) are interesting
- Near-zero returns are very common but uninteresting for traders - unable to cover transaction costs
- Examples:
  - Forecasting a future return of 3% and then it happens -5% is a very bad error!
  - Forecasting a return of 3% and then it happens 11% has the same error amplitude but it is not a serious error
  - Forecasting 0.2% for a true value of 0.4% is reasonably accurate but irrelevant!
  - Forecasting -7.5% for a true value of -8% is a good an useful prediction

DALHOUSIE UNIVERSITY

UBL and Imbalanced Domains

A Side Note

- Frequently, UBL problems are associated with Imbalanced Domains / Datasets
- Actually, imbalanced domains are special cases of UBL
- Properties of Imbalanced Domains:
  - Non-uniform preferences (like UBL)
  - More important cases are rare (not a necessary condition in UBL)

More info at:

P. Branco, L. Torgo and R. Ribeiro. "A Survey of Predictive Modeling on Imbalanced Domains". In: ACM Comput. Surv. 49.2-31 (2016).

Specifying User Preferences

UBL - basics

- UBL is based on domain knowledge - what is relevant
- How to express this information is a key step
  - Different types of information may be available (or missing!)
  - Quality and completeness of domain information may vary
- For classification we typically require a cost/benefit matrix
- What about regression? Infinite domain of the target...
  Thus we need a Utility Surface!
  - A Utility Surface is a function that maps the prediction of value $\hat{y}$ for the true value y into a utility score, $U(\hat{y}, y)$
  - As in classification we assume positive values for correct predictions and negative values for errors
  - Not so obvious what is an error though...

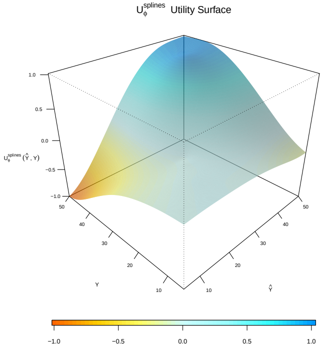
DALHOUSIE UNIVERSITY

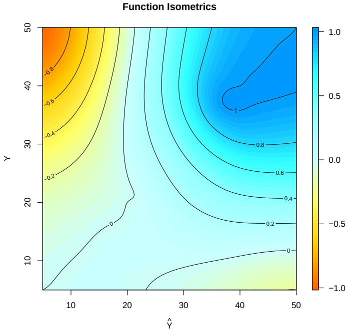
Utility Surfaces - 1

- Asking the user to fully specify a Utility Surface is not reasonable!
- Two options remain:
  1. Use interpolation methods
     - ⋆ General
     - ⋆ Requires the user intervention by providing some points of the surface
  2. Use the automatic method proposed by Ribeiro (2011)
     - ⋆ Only applicable to a special class of UBL problems
     - ⋆ Does not require user intervention

R. Ribeiro (2011). "Utility-based Regression". PhD on Computer Science, Univ. Porto.

Utility Surfaces through Interpolation





- User specifies a few (important) points of the surface
- Some interpolation method is used to obtain the surface
- Method available in the R package UBL (Branco et al., 2016)

P. Branco, R. Ribeiro, L. Torgo (2016). "UBL: an R package for Utility-based Learning". arXiv:1604.08079 [cs.MS].

DALHOUSIE UNIVERSITY

Utility Surfaces through Automatic Derivation - 1

- Ribeiro (2011) proposed a method to automatically defining Utility Surfaces for a particular class of UBL regression applications:
  - Users' goals are rare extreme values of the continuous target
  - This is a **special case of imbalanced regression** domains
- Although a special case, this matches many important applications of utility-based regression
- Branco et al. (2017) have recently shown that a similar approach can also be applied to multi-class imbalance domains

R. Ribeiro (2011). "Utility-based Regression". PhD on Computer Science, Univ. Porto.

P. Branco, L. Torgo and R. P. Ribeiro (2017). "Relevance-based Evaluation Metrics for Multi-class Imbalanced Domains". In: 21th Pacific-Asia Conference, PAKDD'17. Springer.

Utility Surfaces through Automatic Derivation - 2

- Ribeiro's approach is based on the concept of relevance function, $\phi()$, proposed in Torgo and Ribeiro (2007).
- This function is used to specify user-preferences.
- The function maps the continuous domain of the target into a $[0, 1]$ scale of importance

$$\phi(Y) : \mathcal{Y} \rightarrow [0, 1] \tag{1}$$

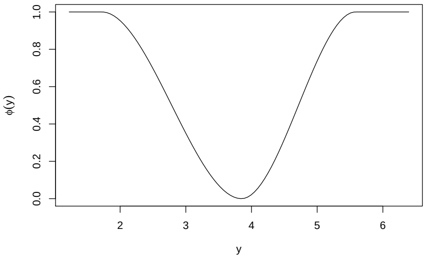
- This function can be automatically derived if assuming rare extremes are the target

L. Torgo and R. Ribeiro (2007). "Utility-based Regression". In: Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2007). Springer.

DALHOUSIE UNIVERSITY

Examples

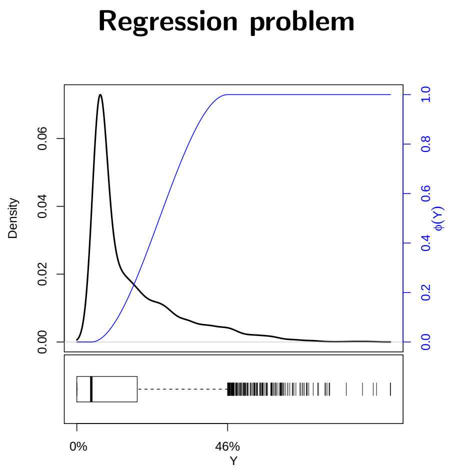
Regression problem

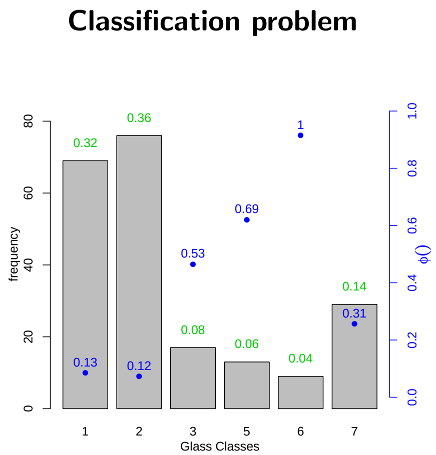


Classification problem

$$\phi(c) = \begin{cases} 0.2 & \text{if } c = normal \\ 0.9 & \text{if } c = rare1 \\ 1 & \text{if } c = rare2 \end{cases}$$

Examples of Automatic Derivation of $\phi()$





DALHOUSIE UNIVERSITY

From Relevance Functions to Utility Surfaces - 1

- The Relevance Function tells us how important is a given value of the target variable to the end-user
- **How to translate that into the cost or benefit of forecasting $\hat{y}$ for a true value of y ?**
- In standard regression this is normally evaluated by $|\hat{y} - y|$ or $(\hat{y} - y)^2$
- But these metrics do not take into account the end-user preferences!
  - Forecasting 10 for a true value of 11.5 will always have the same "value" as forecasting -0.5 for a true value of 1
  - But these may be very different situations for the end user!
- How to use $\phi()$ to decide on the utility of predicting $\hat{y}$ for y, $U(\hat{y}, y)$?

From Relevance Functions to Utility Surfaces - 2

- Torgo and Ribeiro (2007) have proposed to use $\phi()$ to define $U(\hat{y}, y)$ as a net balance between the benefit and the costs of predicting $\hat{y}$ for y
- This definition was based on the proposal of the bi-variate relevance function, $\phi^p(\hat{y}, y)$, that balances the relevance of both $\hat{y}$ and y,

$$\phi^p(\hat{y}, y) = (1 - p) \cdot \phi(\hat{y}) + p \cdot \phi(y)$$

- p defines the relative importance given to false alarms vs missed important values (with $p = 0.5$ giving equal importance)

L. Torgo and R. Ribeiro (2007). "Utility-based Regression". In: Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2007). Springer.

From Relevance Functions to Utility Surfaces - 3

- Ribeiro (2011) has proposed a refined definition of $U(\hat{y}, y)$,

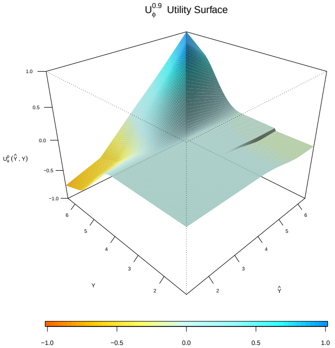
$$\begin{aligned} U_\phi^p(\hat{y}, y) &= B_\phi(\hat{y}, y) \; - \; C_\phi^p(\hat{y}, y) \\ &= \phi(y) \cdot (1 - \Gamma_B(\hat{y}, y)) \; - \; \phi^p(\hat{y}, y) \cdot \Gamma_C(\hat{y}, y) \end{aligned}$$

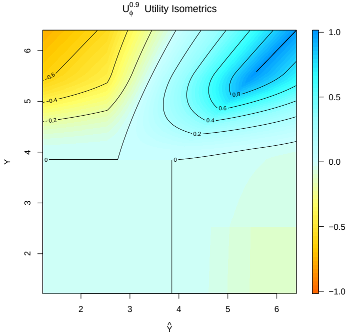
where $\Gamma_B(\hat{y}, y)$ and $\Gamma_C(\hat{y}, y)$ are two bounded-loss functions. Both "normalize" the standard regression loss (squared or absolute) that has domain $[0, \infty]$ to an interval $[0, 1]$

- $\Gamma_B(\hat{y}, y), \Gamma_C(\hat{y}, y) \to 0$ as $\hat{y}$ gets near to y
- $U_\phi^p(\hat{y}, y)$ has domain $[-1, 1]$
- The maximum of $U_\phi^p(\hat{y}, y)$ is $\phi(y)$ when $\hat{y} = y$
- The minimum $U_\phi^p(\hat{y}, y)$ is $p \cdot (1 - \phi(y)) - 1$

R. Ribeiro (2011). "Utility-based Regression". PhD on Computer Science, Univ. Porto.

From Relevance Functions to Utility Surfaces - 4





Performance Evaluation for UBR

Performance Evaluation in Utility-based Regression

- Having defined a Utility Surface we can use it to evaluate a model, e.g. Mean Utility over a test set

$$MU = \frac{1}{N} \sum_{i=1}^{N} U(\hat{y}_i, y_i)$$

Performance Evaluation for Imbalanced Regression

For forecasting rare extreme values, the concepts of Precision and Recall were also adapted to regression (Torgo and Ribeiro, 2009; Branco, 2014),

$$prec^{\phi} = \frac{\sum_{\phi(\hat{y}_i) > t_R} (1 + U(\hat{y}_i, y_i))}{\sum_{\phi(\hat{y}_i) > t_R} (1 + \phi(\hat{y}_i))}$$

$$rec^{\phi} = \frac{\sum_{\phi(y_i) > t_R} (1 + U(\hat{y}_i, y_i))}{\sum_{\phi(y_i) > t_R} (1 + \phi(y_i))}$$

L. Torgo and R. P. Ribeiro (2009). "Precision and Recall for Regression". In: Discovery Science'2009. Springer.

P. Branco (2014). "Re-sampling Approaches for Regression Tasks under Imbalanced Domains". MSc on Computer Science, Univ. Porto.

Learning Models that Maximize Utility

Learning by Optimizing Utility

Formalization

For a given test case $q = \langle \mathbf{x}_k, y_k \rangle$, with y_k unknown, the optimal prediction y^* can be determined as follows:

$$y^* = \underset{z \in Y}{\operatorname{argmax}} \int f_{Y|X}(y|X = \mathbf{x}_k) \cdot U(y, z)\ dy \tag{2}$$

- This equation is an extension of the concept of miminization of conditional risk to a utility-based regression setting
- Having defined $U()$ we only need an estimate of the conditional density probability function, $f_{Y|X}(y|X = \mathbf{x}_k)$

Approximating the Conditional Density Probability Function - 1

- Branco et al. (2017) proposed an approach to approximate $f_{Y|X}(y|X = \mathbf{x}_k)$
- The approach is based on using simpler class density probability function estimators
- Specifically, a method proposed by Frank and Bouckaert (2009) based on ordinal classification

P. Branco, L. Torgo, R. Ribeiro, E. Frank, B. Pfahringer and M. Rau (2017). "Learning Through Utility Optimization in Regression Tasks". In: IEEE International Conference on Data Science and Advanced Analytics (DSAA'17).

E. Frank and R. Bouckaert (2009). "Conditional density estimation withclass probability estimators". in Asian Conference on Machine Learning. Springer.

Approximating the Conditional Density Probability Function - 2

- Assume we have a class probability estimator $\hat{p}$ that given a test case q produces an estimate of the class probability $\hat{p}(c|q)$.
- These estimated probabilities are used as weights of each value of the target conditioned on q

$$w_i(q) = \frac{\hat{p}(c_{y_i}|q)}{n_{c_{y_i}}} \tag{3}$$

- Finally, a weighted kernel Gaussian estimator is used to approximate $f_{Y|X}(y|X = \mathbf{x}_k)$,

$$f_{Y|X}(y|X = \mathbf{x}_k) = \sum_{i=1}^{n} w_i(q)\mathcal{N}(y; y_i, \sigma^2) \tag{4}$$

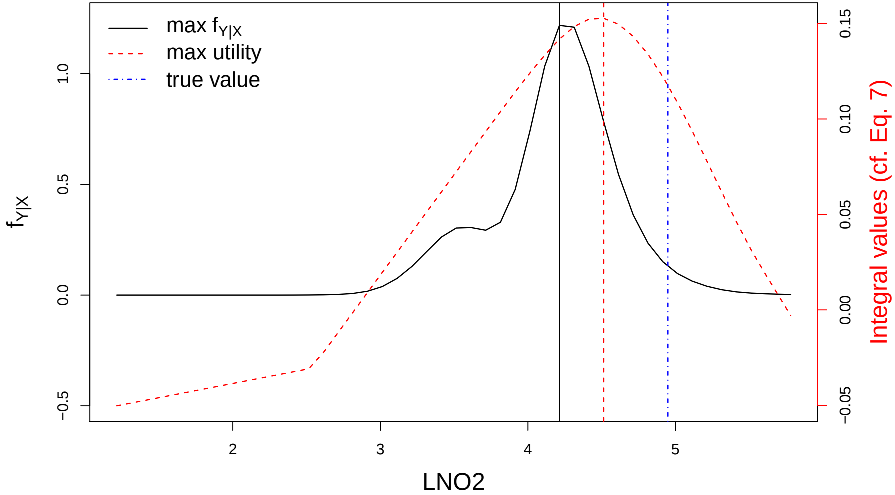
DALHOUSIE UNIVERSITY

Approximating the Conditional Density Probability Function - 3

- The class probabilities $\hat{p}(c|q)$ are obtained by using a classification method
- This method is trained on a discretized version of the training set
- The continuous target variable is discretized into a set of equal width, non-overlapping bins

DALHOUSIE UNIVERSITY

An Example of the Adjustments



Results

- Overall, the utility optimization strategy achieved interesting results, particularly when using SVMs as learners

Figure: Wins (left) and losses (right) of the utility optimization strategy against the baseline for a utility surface with p=0.5.

DALHOUSIE UNIVERSITY

Summary and Challenges

Summary

- UBR is not studied enough in ML and DM communities
- Many relevant applications exist
- Discretizing the target variable and handle the problem as a utility-based classification task is not effective!
- Several solutions have already been proposed for the main challenges
  - Specification of user preferences
  - Evaluation according to these preferences
  - Obtaining models that optimize these preferences

DALHOUSIE UNIVERSITY

Challenges

- Making the ML and DM communities aware of these problems
- Improve current solutions
  - Can we obtain utility surfaces in a easier way?
  - Can we provide new evaluation metrics?
  - What about graphical tools for evaluation?
  - New efforts are needed for learning regression models that optimize utility
    - ★ Better ways of approximating the conditional density probability function
    - ★ Incorporating the utility function in the learning algorithm
  - Provide new solutions and/or guidelines for the related problem of imbalance regression

Utility-based Regression

Luis Torgo, Paula Branco and Rita Ribeiro

Dalhousie Univ., Canada
INESC Tec, Univ. Porto, Portugal

May, 2018