

Classification-relevant Importance Measures for the West German Business Cycle

Daniel Enache, Claus Weihs and Ursula Garczarek

Department of Statistics, University of Dortmund*, 44221 Dortmund, Germany

Abstract. When analyzing business cycle data, one observes that the relevant predictor variables are often highly correlated. This paper presents a method to obtain measures of importance for the classification of data in which such multicollinearity is present. In systems with highly correlated variables it is interesting to know what changes are inflicted when a certain predictor is changed by one unit and all other predictors according to their correlation to the first instead of a *ceteris paribus* analysis. The approach described in this paper uses directional derivatives to obtain such importance measures. It is shown how the interesting directions can be estimated and different evaluation strategies for characteristics of classification models are presented. The method is then applied to linear discriminant analysis and multinomial logit for the classification of west German business cycle phases.

1 Important Measures in Classification

1.1 Economic Multipliers

Usually, the influence of one variable x_i out of a set of variables x_j , $j = 1, \dots, p$, on a function f at a given point \mathbf{x}^0 can be measured by the partial derivative $\partial f(\mathbf{x}^0)/\partial x_i$. This measure can be interpreted as the change of f , if x_i is changed by one unit, given that the other variables x_j , $j = 1, \dots, p$, $i \neq j$, are held constant (*ceteris paribus*).

This information is only of value, if the variables are uncorrelated. An interesting approach to obtain more reliable measures that can be interpreted like economic multipliers and is based upon averaging over orderings, see Lindeman et al. (1980) and Kruskal (1987) for linear regression and Enache and Weihs (2005) for classification.

If multicollinearity is present, the economic multipliers obtained by these methods only reflect the relative importance of the predictors. The effect of changing one predictor variable, e.g. by fiscal policy, on the result is not measured realistically, since the change in one variable inflicts changes in the other variables as well.

1.2 Directional Derivatives

It is interesting to analyze the effects of changing one predictor along with the change of all the correlated variables on the result, for example, if one is interested in the effects of a certain fiscal policy action. To incorporate the relationships to all other variables correctly, directional derivatives instead of partial derivatives have to be used.

* This work has been supported by the Collaborative Research Center ‘Reduction of Complexity in Multivariate Data Structures’ (SFB 475) of the German Research Foundation (DFG).

The directional derivative of function f with respect to $\mathbf{x} \in \mathbb{R}^p$ at point \mathbf{x}_0 in direction $\mathbf{d} \in \mathbb{R}^p$, $\|\mathbf{d}\| = 1$, is defined as:

$$\frac{\partial_{\mathbf{d}} f(\mathbf{x}_0)}{\partial_{\mathbf{d}} \mathbf{x}} = \lim_{t \rightarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{d}) - f(\mathbf{x}_0)}{t}, \quad (1)$$

if the limit exists. For practical purposes, the relation between directional derivatives and the gradient can be used¹:

$$\frac{\partial_{\mathbf{d}} f(\mathbf{x}_0)}{\partial_{\mathbf{d}} \mathbf{x}} = \mathbf{d}' \frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{x}}. \quad (2)$$

1.3 Interpretation of Directional Derivatives

Obviously, partial derivatives are a special case of directional derivatives, that is, derivatives in direction of the unit vectors. With directional derivatives, not only the change of one variable x_i is analyzed, but the simultaneous change of all variables x_1, \dots, x_p , described by the direction in \mathbf{d} . It can be interpreted geometrically as the slope of the tangential hyperplane at point \mathbf{x}_0 , but unlike partial derivatives, the direction in which the slope is measured is not the coordinate axis of x_i but rather the direction described by \mathbf{d} .

1.4 Estimation of the Direction Vectors

In order to develop importance measures based on directional derivatives, the correct direction vectors have to be chosen. The direction vector corresponding to a change of x_j by one unit and the corresponding changes of the other variables according to their correlation with x_j is denoted by \mathbf{d}_j . The single elements of the direction vectors can be estimated using simple linear regressions of all x_i on x_j : $x_i = a_{ij} + b_{ij}x_j + \varepsilon_i$, $i = 1, \dots, p$. The coefficients b_{ij} can be estimated by s_{ij}/s_{jj} and combined to a vector $\tilde{\mathbf{d}}_j$ of estimated slope coefficients which then has to be normalized to unit length, which gives

$$\hat{\mathbf{d}}_j = \frac{\tilde{\mathbf{d}}_j}{\|\tilde{\mathbf{d}}_j\|} = \begin{pmatrix} \frac{s_{1j}}{s_{jj}} \cdot \frac{s_{jj}}{\sqrt{\sum_{i=1}^p s_{ij}^2}} \\ \vdots \\ \frac{s_{pj}}{s_{jj}} \cdot \frac{s_{jj}}{\sqrt{\sum_{i=1}^p s_{ij}^2}} \end{pmatrix} = \begin{pmatrix} \frac{s_{1j}}{\sqrt{\sum_{i=1}^p s_{ij}^2}} \\ \vdots \\ \frac{s_{pj}}{\sqrt{\sum_{i=1}^p s_{ij}^2}} \end{pmatrix} = \frac{\mathbf{S}_{\bullet j}}{\|\mathbf{S}_{\bullet j}\|}, \quad (3)$$

that is, the j th normalized column of the empirical covariance matrix.

1.5 Importance Criteria in Classification

In classification models with K classes the influence of p predictor variables \mathbf{x} on $p(k|\mathbf{x})$, the posterior class probability of class k , is of interest. Since there are K classes, the data has to be split into classes not only to estimate class means and covariance matrices, but also individual class specific direction vectors \mathbf{d}_{ki} for each variable i . The partial derivatives

¹ The prime ' denotes the matrix transpose and $\frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{x}}$ the gradient.

of selected classification models are given in section 2. The directional derivatives can then easily be computed using equation (2). High positive values meaning a strong increase of the posterior class probability due to the change of the predictor and high negative values indicating a strong decrease.

The posterior probability function is usually not linear and therefore the points at which the derivatives are to be evaluated are to be chosen carefully. In classification it is necessary to distinguish different importance criteria, which help choosing the corresponding evaluation strategy².

Importance for class characterization

This importance measure should give an impression about the influence of the respective variable for staying inside the class, if the class membership is already quite obvious. For this purpose, the direction derivative can be evaluated at the respective class mean.

Importance for class separation

- evaluate at the borders between classes and average
- evaluate at uniformly distributed randomly chosen points which are closest to the class borders and average

Such measures should give an impression about which variable has a strong importance for staying inside the class or entering another class.

The first strategy would be more precise, but for some classification methods, the class borders cannot be obtained analytically and have to be searched using grid search algorithms, which are computationally demanding for higher dimensions. The second strategy can then be used as an approximation.

Overall importance

- evaluate at all data points in the sample and average
- evaluate at equally distanced grid points and average
- evaluate at uniformly distributed random points and average

The first strategy takes the characteristics of the data into account, whereas the second strategy tries to cover the space as best as possible. For high dimensional problems, this is probably computationally not feasible and therefore the third strategy is more convenient as an approximation of the second.

² In the classification of times series data, especially cycle data, it may also be interesting, which variables have the most influence on the class change and one could evaluate the derivatives at the data points where these changes occur.

2 Derivatives for the Selected Classification Models

Since classical LDA without dimension reduction is used here, both LDA and Logit have K linear discriminant functions, and can therefore be written as

$$p(k|\mathbf{x}) = \frac{\exp(\alpha_k + \mathbf{x}'\boldsymbol{\beta}_k)}{\sum_{i=1}^K \exp(\alpha_i + \mathbf{x}'\boldsymbol{\beta}_i)}, \quad i = 1, \dots, K, \quad (4)$$

with gradients

$$\frac{\partial p(k|\mathbf{x}_0)}{\partial \mathbf{x}} = p(k|\mathbf{x}_0) \cdot \left[\sum_{i=1}^K p(i|\mathbf{x}_0) \cdot (\boldsymbol{\beta}_k - \boldsymbol{\beta}_i)' \right], \quad k = 1, \dots, K. \quad (5)$$

For LDA the coefficients are $\alpha_k = -\frac{1}{2}\boldsymbol{\mu}'_k \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \ln(p(k))$ and $\boldsymbol{\beta}_k = \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k$. For the Logit model, α_1 and $\boldsymbol{\beta}_1$ are set to 0 (reference class). The unknown parameters for each model can be estimated from the sample.

3 Classification of the West-German Business Cycle

3.1 Data

The application of the importance measures is the classification of business cycle phases. Quarterly west german macro economic data is analyzed, consisting of 157 observations from 4/1955 to 4/1994. There are four classes corresponding to the four phases upswing (“up”), upper turning point (“utp”), downswing (“down”), and lower turning point (“ltp”).

Heilemann and Münch (1996) reduced the predictors to a set of 13 important variables:

- Y yearly growth rate of the gross national product (GNP)
- C yearly growth rate of private consumption
- GD government deficit as a proportion of the GNP
- L yearly growth rate of the number of wage and salary earners
- X netto exports as a proportion of the GNP
- M1 yearly growth rate of money supply
- IE yearly growth rate of equipment investments
- IC yearly growth rate of construction investments
- LC yearly growth rate of labour unit costs
- PY yearly growth rate of GNP price deflator
- PC yearly growth rate of the consumer price index
- RS nominal short term interest rate
- RL real long term interest rate

There is multicollinearity present in the data. The highest positive correlations are 0.868 for LC and PY and 0.776 for C and Y. The highest negative correlations are -0.656 for RL and LC and -0.656 for RL and PY. The variable PY has a variance inflation factor³

³ VIF is the factor, by which the variance of the estimated linear regression coefficient is increased due to multicollinearity. Typically, 10 is considered a critical value, in some references 7.

(VIF, e.g. Neter et al. (1990)) of 16.254, which is definitely above 10. There are also three variables (Y, LC, and RL) with variance inflation factors over 7. VIF's are computed for the whole sample. Within-class multicollinearity is even stronger (e.g. VIF = 275.467 for PY).

3.2 Results

Class Characterization

Table 1 shows the results for both, the linear discriminant analysis and the multinomial logit.

Var	up		utp		down		ltp	
	LDA	Logit	LDA	Logit	LDA	Logit	LDA	Logit
Y	0.0003	-0.0007	0.0350	0.0210	-0.0086	-0.0072	-0.0129	-0.0010
C	-0.0046	-0.0026	0.0332	0.0162	-0.0063	-0.0046	-0.0076	-0.0004
GD	-0.0057	-0.0034	-0.0508	-0.0204	-0.0061	-0.0055	-0.0313	-0.0022
L	0.0012	-0.0021	0.0543	0.0290	-0.0077	-0.0072	-0.0218	-0.0016
X	0.0141	0.0050	0.0399	0.0180	0.0050	0.0093	-0.0036	-0.0008
M1	-0.0031	-0.0014	0.0425	0.0222	-0.0110	-0.0024	-0.0004	0.0000
IE	0.0028	0.0003	0.0267	0.0212	-0.0064	-0.0064	-0.0079	-0.0006
IC	-0.0012	-0.0008	0.0069	0.0019	-0.0069	-0.0028	-0.0109	-0.0008
LC	-0.0120	-0.0066	-0.0681	-0.0364	0.0060	-0.0017	0.0079	0.0009
PY	-0.0112	-0.0056	-0.0527	-0.0260	0.0056	-0.0030	0.0125	0.0011
PC	-0.0197	-0.0049	-0.0158	-0.0120	0.0100	-0.0018	0.0157	0.0012
RS	0.0003	0.0005	0.0374	0.0142	0.0176	0.0037	0.0068	0.0006
RL	0.0058	0.0049	0.0582	0.0272	0.0021	0.0011	-0.0038	-0.0006

Table 1. Variable importances (directional derivatives) for class characterization. The results of classical linear discriminant analysis and multinomial logit.

The three most important variables for class “up” using LDA are PC, X, and LC. The directional derivative for PC is -0.0197. This means if PC is increased by one unit (and all other variables according to their relationship with this variable) and the system is really deep in an upswing phase then the probability of remaining in that phase is decreasing by 1.97 percent points. Using the logit model the variable PY appears to be more important than PC. For the “utp” class the three most important variables are LC, RL, and L in both models. In the “down” class the two models identify different sets of important variables. Whereas in LDA the three most important are RS, M1, and PC, the logit identifies X, Y, and L. Important for the class characterization of the “ltp” class in both models are GD, L, and PC.

Note, that the turning point classes in LDA have higher derivatives than the others, meaning that their class probabilities change easier than those of the upswing and downswing classes. For the logit on the other hand the class “ltp” has low derivatives at the class centers.

Separation

Since a grid search for the class borders is computationally very demanding, points at the borders are searched using a random simulation. (1,000 uniformly distributed data points

have been generated for each class and the 10% points closest to the borders have been chosen.) Table 2 shows the resulting importance measures for class separation.

Var	up		utp		down		ltp	
	LDA	Logit	LDA	Logit	LDA	Logit	LDA	Logit
Y	-0.0002	0.0048	0.0269	0.0231	-0.0126	-0.0010	-0.0134	-0.0097
C	-0.0093	-0.0117	0.0244	0.0170	-0.0097	-0.0015	-0.0092	-0.0066
GD	-0.0097	-0.0182	-0.0331	-0.0151	-0.0087	-0.0020	-0.0267	-0.0300
L	0.0032	0.0059	0.0440	0.0248	-0.0116	0.0004	-0.0211	-0.0252
X	0.0261	0.0404	0.0251	0.0170	0.0076	0.0036	0.0001	0.0038
M1	-0.0073	-0.0172	0.0254	0.0199	-0.0134	-0.0037	-0.0036	-0.0020
IE	0.0041	0.0129	0.0241	0.0128	-0.0091	-0.0014	-0.0097	-0.0048
IC	-0.0022	-0.0043	0.0077	0.0019	-0.0095	-0.0004	-0.0113	-0.0075
LC	-0.0195	-0.0462	-0.0459	-0.0218	0.0057	0.0023	0.0077	-0.0056
PY	-0.0221	-0.0430	-0.0411	-0.0122	0.0051	0.0021	0.0108	-0.0020
PC	-0.0371	-0.0331	-0.0149	-0.0045	0.0117	0.0024	0.0141	0.0025
RS	0.0002	0.0132	0.0190	0.0208	0.0196	0.0060	0.0075	-0.0042
RL	0.0091	0.0304	0.0369	0.0189	0.0030	-0.0004	-0.0022	0.0111

Table 2. Variable importances (directional derivatives) for class separation. The results of classical linear discriminant analysis and multinomial logit.

For class “up” the three most important variables are PC, X, and PY. The value of -0.0371 for PC means that increasing PC by one unit will decrease the probability of staying in the upswing class by 3.7 percent points. In the logit model LC is more important than PC. LC, L, and PY are the three most important variables for class “utp” in the LDA model, whereas in the logit model Y is more important than PY. In the “down” class LDA identifies RS, M1, and Y as the three most important variables, in contrast to the logit where X is more important than Y. For the “ltp” class the most important variables for LDA are GD, L, and PC, with RL instead of PC for the logit.

Overall Importance

Here, the directional derivatives have been evaluated at each point in the sample and the resulting values have been averaged. Table 3 shows the overall importance measures.

Var	up		utp		down		ltp	
	LDA	Logit	LDA	Logit	LDA	Logit	LDA	Logit
Y	0.0025	0.0020	0.0110	0.0124	-0.0049	-0.0043	-0.0077	-0.0076
C	-0.0065	-0.0087	0.0105	0.0092	-0.0034	-0.0029	-0.0049	-0.0040
GD	-0.0097	-0.0138	-0.0158	-0.0093	-0.0048	-0.0050	-0.0200	-0.0239
L	0.0055	-0.0002	0.0168	0.0152	-0.0028	-0.0004	-0.0141	-0.0183
X	0.0274	0.0260	0.0117	0.0092	0.0028	0.0110	-0.0006	-0.0026
M1	-0.0054	-0.0098	0.0127	0.0110	-0.0128	-0.0120	-0.0004	0.0003
IE	0.0075	0.0077	0.0077	0.0092	-0.0044	-0.0047	-0.0044	-0.0035
IC	-0.0018	-0.0033	0.0027	0.0016	-0.0050	-0.0022	-0.0065	-0.0061
LC	-0.0235	-0.0332	-0.0204	-0.0150	0.0096	0.0133	0.0032	-0.0004
PY	-0.0204	-0.0279	-0.0159	-0.0101	0.0091	0.0125	0.0061	0.0029
PC	-0.0372	-0.0220	-0.0042	-0.0041	0.0119	0.0083	0.0084	0.0053
RS	0.0022	0.0076	0.0120	0.0095	0.0232	0.0247	0.0030	-0.0006
RL	0.0106	0.0226	0.0176	0.0119	0.0017	-0.0034	-0.0005	0.0035

Table 3. Overall variable importances (directional derivatives) for each class. The results of classical linear discriminant analysis and multinomial logit.

	LDA	Logit
up	X, LC, PY, PC	X, LC, PY, PC
utp	L, LC, RL	L, LC, RL, Y
down	M1, RS, PC	M1, RS, X
ltp	Y, GD, L	Y, GD, L
overall	L, LC, PC	L, LC, X

Table 4. Summary of important variables over all criteria. If one variable is important in more than one class, it is listed in the “overall” row.

In the “up” class the three most important variables for LDA are PC, X, and LC. A PC value of -0.0372 means that increasing PC by one unit will result in a decrease of the probability for staying in the upswing class. For the logit model RL appears to be more important than PC. The three most important variables for the “utp” class are LC, RL, and L for LDA and L, LC, and Y for the logit. In the “down” class RS, M1, and PC are important for LDA and RS, LC, and PY for the logit. And finally, for the “ltp” class the most important variables are GD, L, and Y in both models.

4 Summary and Outlook

A method has been introduced for obtaining measures of relative importance of correlated predictors on the value change of a dependent variable using directional derivatives. It can be used with every differentiable output function, like linear and non linear regression or probability functions used in supervised classification. For classification problems, several strategies for choosing evaluation points have been presented and have been applied to the classification of west German business cycle phases.

When comparing the importance measures for all criteria (characterization, separation, and overall importance), one can see, that several variables are important for more than one criterion. These important variables are listed in table 4. Apparently, L and LC have the highest importance overall, a result confirming earlier studies (e.g. Weihs and Garczarek (2002)). When using LDA, PC also is considered important and X likewise when using the logit model. It can also be seen, that each class appears to have a different set of important variables. Whereas most of the classes are dominated by production and labor market variables, for “utp” and “down” classes also monetary variables (M1, RS, and RL) have an important influence.

In future work, directional derivatives of other classification methods, like Fisher’s LDA with dimension reduction, Quadratic Discriminant Analysis, and Support Vector Machines, still have to be derived. Other real world applications require quantitative and qualitative variables as predictors. Obtaining the direction vectors for dummy variables used to represent qualitative predictors is a problem which has to be addressed first.

References

- ENACHE, D. and WEIHS, C. (2005): Importance Assessment of Correlated Predictors in Business Cycles Classification. In: C. Weihs and W. Gaul (Eds.): *Classification: The Ubiquitous Challenge*, Springer, Heidelberg, 545–552.
- HEILEMANN, U. and MÜNCH, H. J. (1996): West german business cycles 1963–1994: A multivariate discriminant analysis. In: *CIRET-Conference in Singapore, CIRET-Studien 50*.

- KRUSKAL, W. (1987): Relative importance by averaging over orderings. *The American Statistician*, 41, 6–10.
- LINDEMAN, R. H., MERENDA, P. F., and GOLD, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis*. Scott Foresman, Glenview, IL.
- NETER, J., WASSERMANN, W., and KUNTER, M.H. (1990): *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, 3rd ed., Richard D Irwin.
- WEIHS, C. and GARCZAREK, U. (2002): Stability of multivariate representation of business cycles over time. Sonderforschungsbereich 475, Technical Report 20/2002, University of Dortmund.