Article

Regression analysis on different mitogenic pathways

Shruti Jain
Department of Electronics and Communication Engineering, Jaypee University of Information Technology, Solan-173234, India
E-mail: jain.shruti15@gmail.com

Received 10 February 2016; *Accepted 20 March 2016*; *Published online 1 June 2016*

Abstract
In this paper different regression analysis methods were discussed on three different mitogenic pathways i.e. ERK, MK2 and JNK. Coefficient of determination, ANOVA, T-value, Durban-Watson statistics were calculated for the corresponding three proteins. The model was made using linear modeling using different regression analysis techniques in which different parameters like Mean sq error, Root mean sq error, Mean abs error, Relative sq error, Root relative sq error and Relative abs error were calculated using different analysis like PLS, linear, SVM, random forest etc were calculated. In all respect results with ERK are the best.

Keywords regression analysis; ERK; JNK; MK2.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

MAP kinases are actually a family of protein kinases (Jain, 2012; Janes et al., 2005; Suzanne et al., 2005; Weiss, 2001) that are widely distributed and are found in all eukaryotic organisms. The MAPKs is divided (Jain et al., 2009; Kyriakis et al., 1996; Pearson et al., 2001) into three families as Extracellular-regulated kinase (ERK) (Jain et al., 2010), p38/high osmolarity glycerol (HOG)/Mitogen-activated protein kinase 2 (MK2) (Jain et al., 2010), stress-activated protein kinase (JNK/SAPK) (Jain, et al 2010; Jain, 2014), which leads to cell death/ cell survival. The ERK are activated by differentiation and mitogenic signals while JNK & MK2 are respond to stress and inflammaton.

ERK pathway is activated by the binding of SOS with Grb2 which results in SOS. SOS then activates RAS which leads RAF. RAF activates MEK1 and MEK2 which further activates ERK1 and ERK 2 respectively (Jain et al., 2010). The second most widely studied MAP kinase cascade is the c-Jun NH2-terminal kinase/stress activated protein kinase (JNK/ SAPK) (Brockhaus et al., 1990; Jain, 2014). The JNK are activated when cells are exposed to ultraviolet radiation, heat shock, or inflammatory cytokines. The p38/ MK2 kinase is the well-characterized member of the MAP kinase family. It is activated in response to inflammatory cytokines, endotoxins, and osmotic stress.

In this paper we will study regarding the regression analysis of different MAPK pathways. We have calculated the mean sq error, root mean sq error, mean abs error, relative sq error, root relative sq error and relative abs error using different analysis like PLS, linear, SVM, random forest etc. Coefficient of determination, ANOVA, T-value, Durban-Watson statistics were also calculated for the corresponding three proteins

2 Materials and Methods

In this paper we are using different regression analysis methods on different MAPK pathways i.e., ERK, MK2 and JNK. Ten different concentrations for TNF (Brockhaus et al., 1990; Jain et al., 2011), EGF (Jain, 2015; Libermann et al., 1984; Normanno et al., 2006) and Insulin (Jain, 2011, 2015; Lizcano et al., 2002) were used in ng/ml, i.e., 0-0-0, 5-0-0, 100-0-0, 0-100-0, 5-1-0, 100-100-0, 0-0-500, 0.2-0-1, 5-0-5, 100-0-500.

A) Calculation of coefficient of determination and Durban Watson statistics:

We have calculated the values of coefficient of determination and Durban Watson statistics for three proteins (ERK, MK2 and JNK) which are the main proteins for MAPK pathway. Equation 1 gives the regression coefficient (r^2) equation.

$$r^2 = 1 - \frac{\sum (y_i - f_i)^2}{\sum (y_i - \overline{y})^2} \quad or \quad r^2 = \frac{\sum (f_i - \overline{f})^2}{\sum (y_i - \overline{y})^2} \tag{1}$$

where y_i are the observed values, f_i are the predicted values, $\overline{y}$ are the mean values of observed data and $\overline{f}$ are the mean values of the predicted data. If the regression model is perfect, error sum of square (SSE) is equal to zero, and its r^2 is 1. If the regression model is not perfect, SS_E is equal to SS_T and its r^2 is zero. Inputs are ten different concentrations of TNF- EGF - Insulin (0-0-0, 5-0-0, 100-0-0, 0-100-0, 5-1-0, 100-100-0, 0-0-500, 0.2-0-1, 5-0-5, 100-0-500) All are ng/ml. Outputs are ERK or JNK or MK2 values.

For our data sets of ten concentrations of TNF/EGF/ Insulin for ERK pathway values are as: S = 0.005931, regression coefficient (r^2) = 92.2%, $r^2_{(adj)}$ = 92.0%, $r^2_{(pred)}$ = 91.6%, PRESS = 0.010980, Durbin-Watson Statistics =2.04.

For our data sets of ten concentrations of TNF/EGF/ Insulin for MK2 pathway values are as: S = 0.006480, r^2 = 90.7% , $r^2_{(adj)}$ = 90.4%, $r^2_{(pred)}$ = 90.01%, PRESS = 0.013055, Durbin Watson Statistics = 2.09.

For our data sets of ten concentrations of TNF/EGF/ Insulin for JNK are as: S = 0.01042, (r^2) = 76.0%, adjusted regression coefficient ($r^2_{(adj)}$) = 75.2%, $r^2_{(pred)}$ = 74.19%, PRESS = 0.033734, Durbin Watson Statistics = 1.61.

B) Calculation of Analysis of Variance (ANOVA) (Table 1, 2 and 3):

Table 1 Analysis of Variance (ANOVA) for all combinations of ERK.

Source	**df**	**SS**	**MS**	**F**
Regression	10	0.120516 (SSa)	0.012052 (MSa)	342.60
Residual Error	289	0.010166 (SSe)	0.000035 (MSe)	
Total	299	0.130683		

Table 2 Analysis of Variance (ANOVA) for all combinations of MK2.

Source	**df**	**SS**	**MS**	**F**
Regression	10	0.118548 (SSa)	0.011855 (MSa)	282.33
Residual Error	289	0.012135 (SSe)	0.000042 (MSe)	
Total	299	0.130683		

Table 3 Analysis of Variance (ANOVA) for all combinations of JNK.

Source	**df**	**SS**	**MS**	**F**
Regression	10	0.099332 (SSa)	0.009933 (MSa)	91.57
Residual Error	289	0.031350 (SSe)	0.000108 (MSe)	
Total	299	0.130682		

Table 1, Table 2 and Table 3 shows the mean squares of the regression, sum of squares and residual error for ERK, MK2 & JNK respectively. Table 4, Table 5 & Table 6 shows the standard error coefficient value, *T*-value, *p*- value, *f*-value which was also calculated for ERK, MK2 & JNK respectively. In these tables, SS is the sum of square, df is the degree of freedom, MS is the mean square, SS_a is the sum of square among groups, SS_e is the error sum of square, MS_a is the average variability among groups, MS_e is the average variability within groups. They are calculated as: $MS_a = SS_a/\text{df}$, $MS_e = SS_e/\text{df}$, $F = MS_a/MS_e$

Table 4 Regression analysis in terms of standard error coefficients, *p* value, *T*-value, *F*- value for ERK.

Effect	**Coefficient**	**Standard Error Coefficient**	**t-value**	**p- value**	**VIF (variance inflation factor)**
Constant	0.34684	0.02003	17.32	0.000	
0-0-0	0.00013307	0.00005417	2.46	0.015	35.0
5-0-0	0.00014586	0.00007029	2.08	0.039	8.7
100-0-0	0.00020099	0.00007236	2.78	0.006	97.1
0-100-0	0.00009306	0.00005560	1.67	0.095	21.4
5-1-0	0.00001065	0.00006060	0.18	0.861	10.6
100-100-0	0.00001734	0.00006432	0.27	0.788	77.3
0-0-500	0.00010723	0.00007165	1.50	0.136	77.4
0.2-0-1	0.00009634	0. 00004255	2.26	0.024	129.5
5-0-5	0.00004758	0.00001973	2.41	0.016	247.3
100-0-500	0.00001964	0.00004623	0.42	0.671	168.9

Table 5 Regression analysis in terms of standard error coefficients, *p* value, *T*-value, *F*- value for MK2.

Effect	Coefficient	Standard Error Coefficient	t-value	p- value	VIF (variance inflation factor)
Constant	0.59869	0.03739	16.01	0.000	
0-0-0	0.00002453	0.00005234	0.47	0.640	1.7
5-0-0	-0.00004073	0.00005098	-0.80	0.425	2.1
100-0-0	-0.00033461	0.00003862	-8.67	0.000	13.7
0-100-0	0.00000636	0.00004956	0.13	0.898	2.0
5-1-0	0.00010052	0.00005540	1.81	0.071	4.1
100-100-0	-0.00001375	0.00004750	-0.29	0.772	23.2
0-0-500	0.00003543	0.00005259	0.67	0.501	15.5
0.2-0-1	0.00009931	0.00004462	2.23	0.027	9.9
5-0-5	-0.00027582	0.00004681	-5.89	0.000	18.9
100-0-500	-0.00007383	0.00004397	-1.68	0.094	33.0

Table 6 Regression analysis in terms of standard error coefficients, *p* value, *T*-value, *F*- value for JNK.

Effect	Coefficient	Standard Error Coeff	T-Value	p- Value	F- Value
0-0-0	0.92309	0.04895	18.86	0.000	
5-0-0	0.00012979	0.00008235	1.58	0.116	1.6
100-0-0	-0.00023739	0.00008791	-2.70	0.007	1.5
0-100-0	-0.00075872	0.00006467	-11.73	0.000	1.6
5-1-0	-0.00000557	0.00008274	-0.07	0.946	1.0
100-100-0	-0.00038617	0.00007818	-4.94	0.000	1.8
0-0-500	0.00009802	0.00008049	1.22	0.224	1.2
0.2-0-1	0.00004452	0.00008060	0.55	0.581	1.0
5-0-5	-0.00042354	0.00008249	-5.13	0.000	4.2
100-0-500	-0.00004693	0.00008336	-0.56	0.574	1.6

We clubbed all the concentrations of TNF, EGF and Insulin and only normalized output (ERK, MK2, JNK) were taken and we get the regression equation as:

Final Output for ERK = 0.347 +0.000133 a +0.000146 b +0.000201 c +0.000093 d +0.000011 e +0.000017 f +0.000107 g +0.000096 h +0.000048 i +0.000020 j (2)

Final Output for MK2 = 0.599 +0.000025 a-0.000041 b -0.000335 c +0.000006 d +0.000101 e -0.000014 f +0.000035 g +0.000099 h -0.000276 i -0.000074 j (3)

Final Output for JNK =0.923 + 0.000130a - 0.000237b- 0.000759c - 0.000006d- 0.000386e + 0.000098f +0.000045g-0.000424h-0.000047i -0.000154j. (4)

where a, b, c are the different concentrations of TNF, EGF and Insulin.

We have calculated mean sq error (MSE), root mean sq error (RMSE), mean abs error (MAE), relative sq error (RSE), root relative sq error (RRSE) and relative abs error (RAE) for ERK, MK2 and JNK using different regression analysis like PLS, linear, SVM, KNN, random forest, regression etc was given in Table 7, Table 8 and Table 9 respectively.

- Mean square error: where fi is the predicted value and yi is the actual/ observed value

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 \tag{5}$$

- Root mean sq error: RMSE is used to measure the error rate of a regression model. However, it can only be compared between models whose errors are measured in the same units.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2} \tag{6}$$

- Mean abs error: MAE has the same unit as the original data, and it can only be compared between models whose errors are measured in the same units. It is usually similar in magnitude to RMSE, but slightly smaller.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_i| \tag{7}$$

- Relative sq error: RSE can be compared between models whose errors are measured in the different units.

$$RSE = \frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(\bar{y} - f_i)^2} \tag{8}$$

- Root relative sq error:

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - f_i)^2}{\sum_{i=1}^{n}(\bar{y} - f_i)^2}} \tag{9}$$

- Relative abs error: RAE can be compared between models whose errors are measured in the different units.

$$RAE = \frac{\sum_{i=1}^{n}|y_i - f_i|}{\sum_{i=1}^{n}|\bar{y} - f_i|} \tag{10}$$

Table 7 Various parameters using different regression methods for ERK.

	MSE	RMSE	MAE	RSE	RRSE	RAE
PLS Regression	0.0000	0.0060	0.0048	0.0836	0.2891	0.2551
Linear Regression	0.0000	0.0060	0.0048	0.0836	0.2891	0.2551
SVM Regression	0.0005	0.0216	0.0207	1.0755	1.0371	1.0969
K nearest neighbours regression	0.0001	0.0072	0.0058	0.1189	0.3448	0.3089
Mean	0.0004	0.0210	0.0190	1.0125	1.0062	1.0042
Random Forest regression	0.0000	0.0065	0.0052	0.0964	0.3105	0.2727
Regression tree	0.0000	0.0060	0.0048	0.0840	0.2898	0.2557

Table 8 Various analysis parameters using diff regression methods for MK2.

	MSE	RMSE	MAE	RSE	RRSE	RAE
PLS Regression	0.0000	0.0067	0.0054	0.1017	0.3190	0.2847
Linear Regression	0.0000	0.0067	0.0054	0.1017	0.3190	0.2847
SVM Regression	0.0005	0.0216	0.0207	1.0755	1.0371	1.0969
K nearest neighbours regression	0.0001	0.0072	0.0058	0.1187	0.3445	0.3058
Mean	0.0004	0.0210	0.0190	1.0125	1.0062	1.0042
Random Forest regression	0.0000	0.0070	0.0055	0.1113	0.3336	0.2888
Regression tree	0.0000	0.0060	0.0048	0.0840	0.2898	0.2557

Table 9 Various analysis parameters using diff regression methods for JNK.

	MSE	RMSE	MAE	RSE	RRSE	RAE
PLS Regression	0.0001	0.0106	0.0085	0.2599	0.5098	0.4521
Linear Regression	0.0001	0.0106	0.0085	0.2599	0.5098	0.4521
SVM Regression	0.0005	0.0216	0.0207	1.0755	1.0371	1.0969
K nearest neighbors regression	0.0001	0.0093	0.0066	0.2004	0.4477	0.3517
Mean	0.0004	0.0210	0.0190	1.0125	1.0062	1.0042
Random Forest regression	0.0001	0.0088	0.0067	0.1763	0.4199	0.3552
Regression tree	0.0000	0.0069	0.0051	0.1104	0.3323	0.2692

3 Conclusion

In this paper we have discussed the coefficient of determination, ANOVA, T-value, Durban Watson statistics, P-value, F-value, standard error coefficients for different mitogenic pathways, i.e., JNK, ERK, and MK2. We have also calculated MSE, RMSE, MAE, RSE, etc., using different analysis methods like PLS, linear, SVM, random forest etc. In all respect results with ERK are the best. In future we will find the pdf all proteins using different distribution functions.

References

Brockhaus M, Schoenfeld HJ, Schlaeger EJ, Hunziker W, Lesslauer W, Loetscher H. 1990. Identification of two types of tumor necrosis factor receptors on human cell lines by monoclonal antibodies. Proceedings of the National Academy of Sciences USA, 87: 3127-3131

Jain S, Naik PK, Sharma R. 2009. A computational modeling of cell survival/ death using VHDL and MATLAB simulator. Digest Journal of Nanomaterials and Biostructures (DJNB), 4(4): 863- 879

Jain S, Bhooshan SV, Naik PK. 2010. Model of mitogen activated protein kinases for cell survival/death and its equivalent bio-circuit. Current Research Journal of Biological Sciences, 2(1): 59-71

Jain S, Bhooshan SV, Naik PK. 2011, Mathematical modeling deciphering balance between cell survival and cell death using insulin. Network Biology, 1(1): 46-58

Jain S, Bhooshan SV, Naik PK. 2011. Mathematical modeling deciphering balance between cell survival and cell death using tumor necrosis factor α. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 2(3): 574-583

Jain S. 2012. Communication of signals and responses leading to cell survival / cell death using Engineered Regulatory Networks. PhD Thesis, Jaypee University of Information Technology, Solan, Himachal Pradesh, India

Jain S. 2014. Implementation of fuzzy system using operational transconductance amplifier for ERK pathway of EGF/ Insulin leading to cell survival/ death. Journal of Pharmaceutical and Biomedical Sciences, 4(8): 701-707

Jain S, Chauhan DS. 2015. Mathematical analysis of receptors for survival proteins. International Journal of Pharmaceutical and Biomedical Sciences, 6(3): 164-176

Jain S, Chauhan DS. 2015. Implementation of fuzzy system using different voltages of OTA for JNK pathway leading to cell survival/ death. Network Biology, 5(2): 62-70

Janes KA, John AG, Suzanne G, Peter SK, Douglas LA, Michael YB. 2005. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis; Science, 310: 1646-1653

Kyriakis JM, Avruch J. 1996. Sounding the alarm: protein kinase cascades activated by stress and inflammation. Journal of Biological Chemistry, 271: 24313-24316

Libermann TA , Razon TA., Bartal AD, Yarden Y., Schlessinger J, Soreq H. 1984. Expression of epidermal growth factor receptors in human brain tumors Cancer Research 44:753-760

Lizcano JM, Alessi DR. 2002. The insulin signalling pathway. Current Biology, 12: 236-238

Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR,, Carotenuto A, De Feo G, Caponiqro F, Salomon DS. 2006. Epidermal growth factor receptor (EGFR) signaling in cancer. Gene, 366: 2-16

Pearson G, Robinson F, Beers GT, Xu BE, Karandikar M, Berman K, Cobb MH. 2001. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. Endocrine Reviews, 22(2): 153-183

Suzanne G, Janes KA, John AG, Emily PA, Douglas LA, Peter SK. 2005. A compendium of signals and responses trigerred by prodeath and prosurvival cytokines. Manuscript M500158-MCP200

Weiss R. 2001. Cellular computation and communications using engineered genetic regulatory networks. PhD Thesis, MIT, USA