eit Manufacturing

Aalto University
School of Engineering

Fitting models to data | 24.1.2022

Esko Niemi | Helsinki

EIT Manufacturing is supported by the EIT,
a body of the European Union

TF KnowNet

1

Outline

TF KnowNet

Linear regression

- Introduction
- Simple linear regression – Introductory example
- Multiple linear regression – Fitting the model
- Multiple linear regression – Interpretation of meaning
- Multiple linear regression – Significance
- Multiple linear regression – Assumptions and limitations

Neural networks

eit Manufacturing

Aalto University
School of Engineering

2

2

Regression models

TF KnowNet

- Regression models are used to find and model a statistical dependence between variables. This is done by fitting a function to collected input (independent) and output (dependent) data
- The type of function is chosen based on the type of assumed dependence
- The values of the parameters of the regression function are defined so that some fitting criterion is optimized, usually it is the sum of squares of distances between measured output data and values predicted by the model (Least Squares Method)

Linear regression

- Useful when there is reason to believe that dependencies between inputs and outputs are linear
- Does not explain causality. On the other hand, describes well reality
- Quality of dependence can be evaluated and understanding of the process in question increases
- Very useful in many kinds of production related applications: cost estimation, work content estimation, modeling of various production processes, market demand prediction etc.

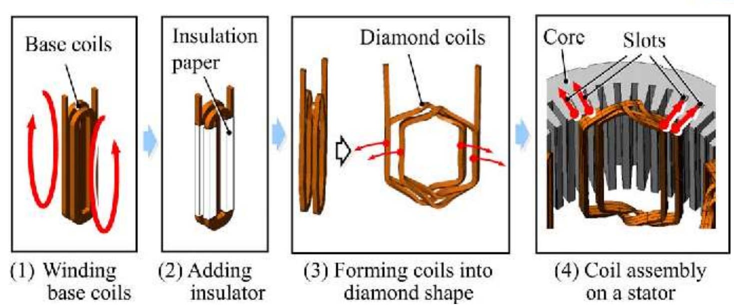
eit Manufacturing

Aalto University
School of Engineering

Linear regression – Stator example

TF KnowNet

We predict work content of manufacturing of stator coils of large electrical AC motors based on delivered orders. Technical specifications of orders are known and working hours have been recorded.



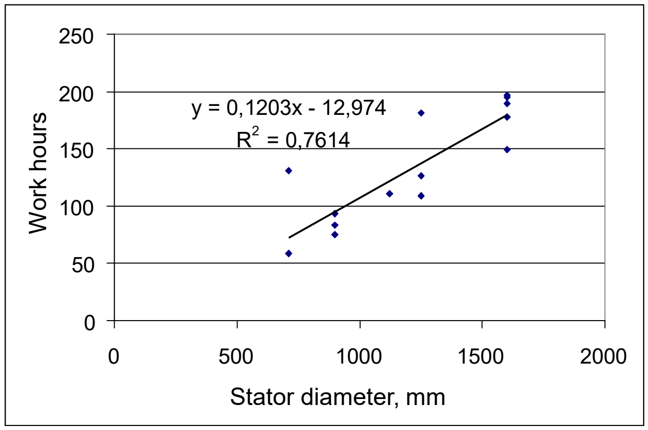
https://www.semanticscholar.org/paper/Motor-Stator-With-Thick-Rectangular-Wire-Lap-for-Ishigami-Tanaka

eit Manufacturing

Aalto University
School of Engineering

Simple linear regression – Stator example

Regression based on one property: We assume that the stator diameter affects the number or size of the coils and therefore the working hours needed to make them.



5

5

Multiple linear regression – Stator example continues

Many factors affect the amount of work required. We want to improve our model and assume that the following are important: Stator diameter (size), stator length (size), number of poles (number of coils in a stator) and voltage (insulation layer thickness). All of them can be taken into account in multiple linear regression analysis.

Linear model is of form:

$$y = b_n x_n + b_{n-1} x_{n-1} + b_{n-2} x_{n-2} + \ldots + b_1 x_1 + b_0$$

Here $x_1 \ldots x_n$ are input variables and $b_0 \ldots b_n$ are constants, the value of which must be determined so that the model predicts as well as possible.

6

6

Multiple linear regression – Stator example continues

TF KnowNET

Stator coil manufacturing working hours

Hours	Stator diameter	Stator length	Number of poles	Voltage	Predicted hours	Error
					Four variable model error	
59	710	650	10	4000	48.4	10.6
83	900	1150	10	4000	90.0	-7.0
111	1120	850	12	5000	105.0	6.0
181	1250	1250	10	13000	166.0	15.0
195	1600	1050	14	13000	197.8	-2.8
126	1250	1350	10	5000	136.9	-10.9
149	1600	750	12	9000	162.7	-13.7
93	900	950	10	11000	110.2	-17.2
178	1600	1050	12	9000	176.9	1.1
190	1600	1350	14	800		
109	1250	750	12	400		
75	900	950	10	300		
197	1600	1250	12	1100		
178	1600	1050	12	600		
131	710	1350	10	1500		
59	710	750	6	700		

Left: Data and calculated hours. Note that R^2 is now 0.967!

Regression Statistics		*Formulas*
Multiple R	0.98319774	0.98319774
R Square	0.966677796	0.966677796
Adjusted R Square	0.954560631	0.954560631
Standard Error	10.63204001	10.63204001
Observations	16	

ANOVA

	df	*SS*	*MS*	*F*
Regression	4	36072.3	9018.1	79.77755504
Residual	11	1243.4	113.0	
Total	15	37315.8		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-86.9467944	19.1529192	-4.53961057	0.000844404
Stator diameter	0.094769015	0.012342802	7.678079546	9.63187E-06
Stator length	0.047320526	0.013166316	3.594059789	0.004213544
Number of poles	2.036548752	2.280536175	0.893013132	0.390968525
Voltage	0.004227445	0.000813669	5.195531958	0.000296524

Right: Results of fitting model using Excel tools (Tools -> Data Analysis -> Regression)

eit Manufacturing

Aalto University
School of Engineering

7

7

Multiple linear regression

TF KnowNET

We fit the multidimensional line equation to our data set by determining the values of b_i so that we minimize

$$SSE = \text{Sum of Squares of Errors} = \Sigma((Y_i - \text{Est}Y_i)^2),$$

where

Y_i = measured data

$\text{Est}Y_i$ = values predicted (calculated) using model

Coefficient of determination, R^2 expresses how big a share of variation in data is explained by the model:

$$R^2 = (1 - (SSE/SST)),$$

where

$$SST = \text{Total Sum of Squares} = \Sigma((Y_i - \text{Mean}Y)^2),$$

where

$\text{Mean}Y$ = average of data point values

eit Manufacturing

Aalto University
School of Engineering

8

8

Multiple linear regression – Interpretation of model

Coefficients b_i of the model determine how much a change in the value of variable x_i affects the result. This effect can be evaluated by:

- Calculating the real contribution. This is done by multiplying average of the related input data values with the coefficient. The result is the expected contribution of each input factor to the average result.
- Standardizing the model. Average of the data values of each type is subtracted from the individual values and divided by the standard deviation of each type of input values. We obtain values, the average of which is 0 and standard deviation 1. By fitting the model to this standardized data the obtained coefficients are comparable and give the relative contribution of the different inputs to the output values. In other words: we can directly conclude which factors affect the result most and are most important

9

9

Multiple linear regression - Significance

SEE = Standard Error of Estimate = Sqrt(SSE/df),

where

df = degrees of freedom = $n - k - 1$,

where

n = number of measurements

k = number of input variables

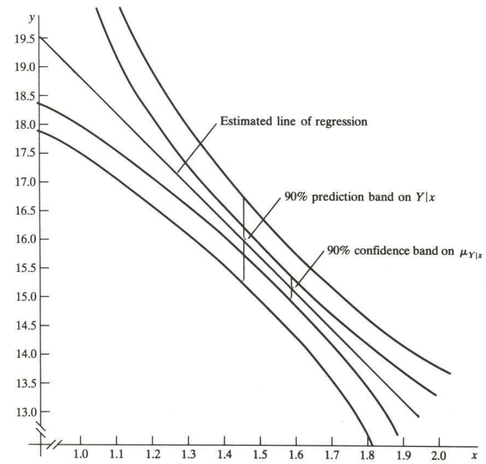
SEE is used for calculating significance of model and its parameters and confidence interval of parameter values and confidence interval of prediction for individual values and the average.

10

10

Multiple linear regression - Significance

Simple linear regression model and confidence intervals for average and a single measurement



Note! You do not get these from Excel's analysis tools, but you get them from statistical software packages like Statistix

11

11

Multiple linear regression - Significance

Significance of R and R^2 and the model can be tested with F test. F depends on R^2, numbers of measurements and variables:

$$F_{k,\, n-k-1} = (\mathrm{Est}Y_i^2/k) \,/\, SEE = [R^2/k] \,/\, [(1 - R^2)/(n - k - 1)].$$

Significance testing is important, because with a large number of variables the R^2 can be high, although SEE is high. In this case the model may not be good with other data than that it was fitted to. The same purpose is served by:

$$\text{Adjusted } R^2 = 1 - ((1-R^2)(N-1 \,/\, N - k - 1)).$$

Adjusted R^2 approaches R^2 when k decreases.

12

12

Multiple linear regression - Significance

Significance of R and R^2 and the model can be tested with F test. F depends on R^2, numbers of measurements and variables:

$$F_{k,\ n-k-1} = (\mathrm{Est}Y_i^2/k) \,/\, SEE = [R^2/k] \,/\, [(1 - R^2)/(n - k - 1)].$$

Significance testing is important, because with a large number of variables the R^2 can be high, although SEE is high. In this case the model may not be good with other data than that it was fitted to. The same purpose is served by:

$$\text{Adjusted } R^2 = 1 - ((1-R^2)(N-1 \,/\, N - k - 1)).$$

Adjusted R^2 approaches R^2 when k decreases.

Multiple linear regression - Assumptions and limitations

- Data samples should represent well the typical behaviour of the phenomenon modelled. The input values (and outputs) should cover the whole typical value range of each variable and typical combinations of input values should appear in the data.
- The model should only contain the factors that truly contribute to the result. Procedures for eliminating unnecessary variables exist and are implemented is statistics software. The number of variables affects the required size of data sample and generally it should be at least 10 – 20 times the number of variables.
- The modelled phenomenon should behave linearly. Linearity can be evaluated by graphing the errors as functions of each variable separately. No distinguishable trend should exist.

Multiple linear regression - Assumptions and limitations

TF KnowNET

- Input variables should be independent of each other. Dependence is called multicollinearity and it can be tested by doing regression analysis of each (x_i) variable's dependence on all the other variables. The R^2 obtained should be low (< 0,2). Multicollinearity shows as large standard error of coefficients (b_i) and low reliability of the model.
- Errors should ne normally distributed. This can be evaluated by drawing a histogram for each variable's errors.
- Regression analysis only describes correlations between variables. It does not explain the reasons for correlations. For example, insurance compensation for losses due to fire correlate strongly with the number of firefighters at the location of fire => ?

eit Manufacturing

Aalto University School of Engineering

15

15

Outline

TF KnowNET

Neural networks

- Introduction
- Neuron
- Network
- Training a single linear neuron
- Nonlinear transfer functions
- Scaling output
- Modeling and using a neural network
- Example application: Estimating cost of anodizing process
- Conclusions

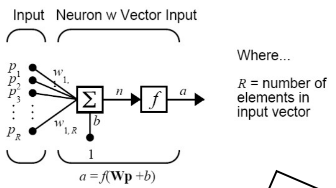
eit Manufacturing

Aalto University School of Engineering

16

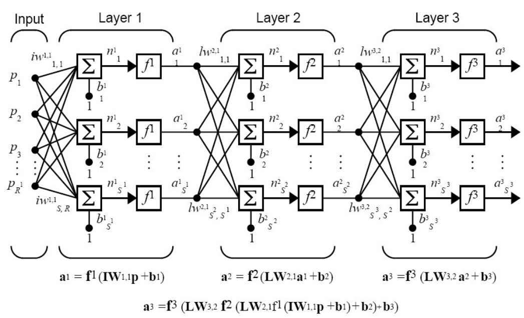
16

Neural networks



Machine learning, an important category of *Artificial intelligence*, is usually realized with *Neural networks*

Most of the figures in this presentation are from *Matlab* documentation



eit Manufacturing

Aalto University School of Engineering

17

17

Neural networks

- Neural network is a ”black box” between given inputs and obtained outputs in a similar fashion to a regression model
- Realization of a neural network and setting of parameters, however, is different, and
  - Neural networks can be used to model nonlinear systems
  - Actually, almost no limiting assumptions about the modeled system’s behavior need to be made
  - Several outputs can be modelled at the same time unlike (most) regression models
- On the other hand, similar justified conclusions about reliability of neural networks can not be done as about linear regression models. They are “unpredictable”.
- Neural networks can be used for classification and clustering with or without training, but here we only focus on process modeling using static *feedforward* networks

eit Manufacturing

Aalto University School of Engineering

18

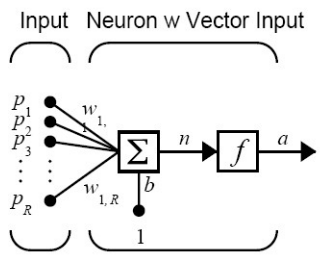
18

Neuron

TF KnowNet

Neural network consists of neurons, which receive inputs (p_i), weigh the inputs with weights (w_i), possibly add a constant, bias (b), calculate the sum (n) and change it with a transfer function *($a = f(n)$)*.

One neuron:



$a = f(\mathbf{Wp} + b)$

Where...

R = number of elements in input vector

eit Manufacturing

Aalto University School of Engineering

19

19

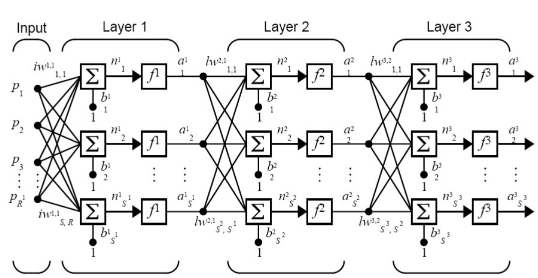
Network of neurons

TF KnowNet

Neurons can be combined with outputs and inputs to form (multiple) layers:

- A neuron can take several inputs
- A neuron has only one output,
- but this output can be connected as input to several neurons
- The last (output) layer has to have as many neurons as there are final outputs
- Layers inside the network are called hidden layers

Input layer only distributes inputs to first neuron layer



$\mathbf{a}^1 = \mathbf{f}^1(\mathbf{IW}^{1,1}\mathbf{p} + \mathbf{b}^1)$ $\quad \mathbf{a}^2 = \mathbf{f}^2(\mathbf{LW}^{2,1}\mathbf{a}^1 + \mathbf{b}^2)$ $\quad \mathbf{a}^3 = \mathbf{f}^3(\mathbf{LW}^{3,2}\mathbf{a}^2 + \mathbf{b}^3)$

$\mathbf{a}^3 = \mathbf{f}^3(\mathbf{LW}^{3,2}\,\mathbf{f}^2(\mathbf{LW}^{2,1}\mathbf{f}^1(\mathbf{IW}^{1,1}\mathbf{p} + \mathbf{b}^1) + \mathbf{b}^2) + \mathbf{b}^3)$

eit Manufactur

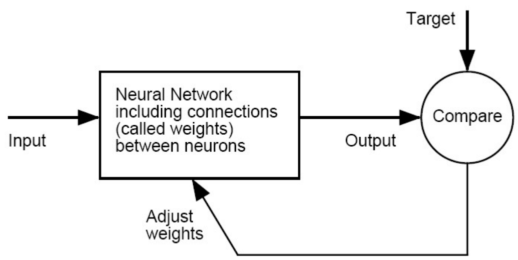
Aalto University School of Engineering

20

20

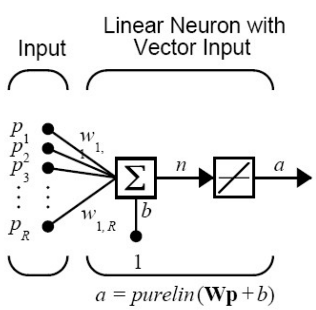
Fitting network to data - training

Network is trained with input-target pairs so, that the output is compared to target and the weights and bias are adjusted according to error and learning algorithm recursively and repeatedly. This is continued as long as error converges or some preset target value for error is reached. The best known learning algorithm is called backpropagation.



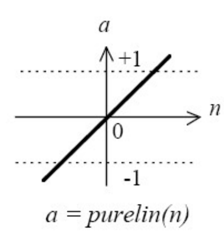
Linear neuron

Clearly, one neuron with a linear transfer function (that does nothing), functions like a linear regression model



Where...

R = number of elements in input vector



Linear Transfer Function

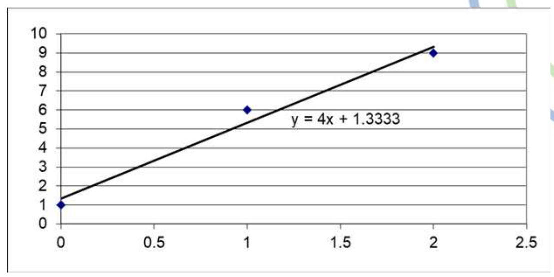
$$n = w_{1,1}p_1 + w_{1,2}p_2 + \ldots + w_{1,R}p_R + b$$

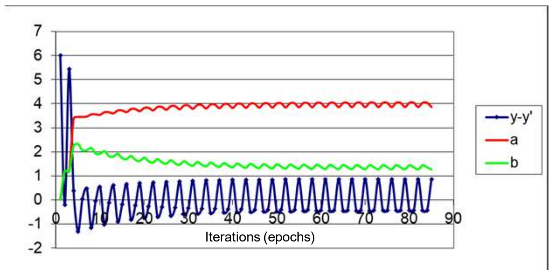
Training a linear neuron - example

C3 = =C2+H2

	A	B	C	D	E	F	G	H	I	J
1	x	y	a	b	y'=ax+b	y-y'	alpha	da	db	SSE
2	1	6	0,000	0,000	0,000	6,000	0,100	0,600	0,6	
3	0	1	0,600	0,600	0,600	0,400	0,100	0,000	0,04	
4	2	9	0,600	0,640	1,840	7,160	0,100	1,432	0,716	87,426
5	1	6	2,032	1,356	3,388	2,612	0,100	0,261	0,2612	
6	0	1	2,293	1,617	1,617	-0,617	0,100	0,000	-0,06172	
7	2	9	2,293	1,555	6,142	2,858	0,100	0,572	0,285812	15,372
8	1	6	2,865	1,841	4,706	1,294	0,100	0,129	0,129388	
9	0	1	2,994	1,971	1,971	-0,971	0,100	0,000	-0,09707	
10	2	9	2,994	1,874	7,862	1,138	0,100	0,228	0,113796	3,911
11	1	6	3,222	1,987	5,209	0,791	0,100	0,079	0,079079	
12	0	1	3,301	2,066	2,066	-1,066	0,100	0,000	-0,10665	
13	2	9	3,301	1,960	8,562	0,438	0,100	0,088	0,043839	1,955
14	1	6	3,389	2,004	5,392	0,608	0,100	0,061	0,060776	
15	0	1	3,449	2,064	2,064	-1,064	0,100	0,000	-0,10645	
16	2	9	3,449	1,958	8,857	0,143	0,100	0,029	0,014331	1,523
17	1	6	3,478	1,972	5,450	0,550	0,100	0,055	0,054966	
18	0	1	3,533	2,027	2,027	-1,027	0,100	0,000	-0,10273	
19	2	9	3,533	1,925	8,991	0,009	0,100	0,002	0,000949	1,358
20	1	6	3,535	1,926	5,460	0,540	0,100	0,054	0,053961	
21	0	1	3,589	1,979	1,979	-0,979	0,100	0,000	-0,09795	

Training data set





eit Manufacturing

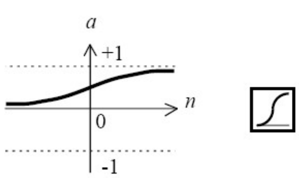
Aalto University School of Engineering

23

23

Nonlinear transfer functions

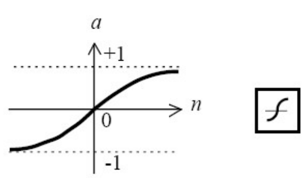
For modeling nonlinear systems transfer functions have to be nonlinear. Sigmoid functions are a typical choice:



$a = logsig(n)$

Log-Sigmoid Transfer Function

$a = 1/(1 + e^{-n})$



$a = tansig(n)$

Tan-Sigmoid Transfer Function

$a = (1 - e^{-2n})/(1 + e^{-2n})$

eit Manufacturing

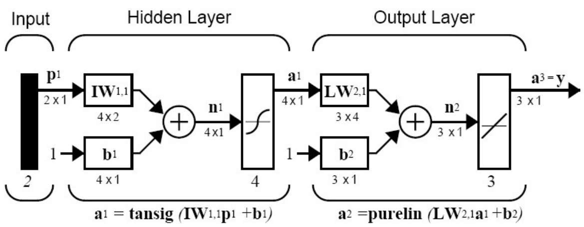
Aalto University School of Engineering

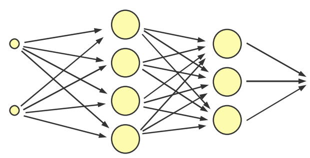
24

24

Scaling output

Because sigmoid functions only return values in range (0, 1) or (-1, 1), output layer may have a linear transfer function, which scales the results:





Neural network modeling process

1. Define purpose, inputs and outputs and collect data.
2. Network design and creation (neuron types, parallel neurons and layers…).
3. Network *training*. Training requires several learning cycles (epochs), during which the weights are adjusted. This may lead to a situation, in which it “memorizes” the training data, but does not work well with other data from the same system. This is common if the modeled process is highly nonlinear or otherwise erratic or the data is not representative. It is essential to *validate* and *test* the network also with other than the training data.
4. Training can be automated as follows: Data is divided to training, validation and test sets. Only training set is used for adjusting weights, but during the process the network is tested with the validation set. When fit to the validation set does not improve anymore, training is stopped. This method aims at ensuring network’s generalization capability.
5. Network is tested with the test set and used.

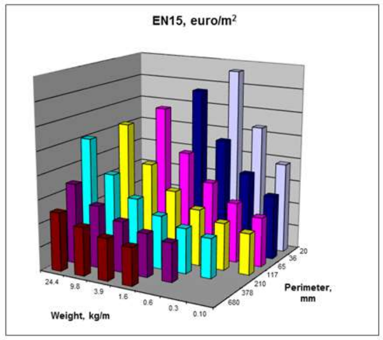
Example: cost of anodizing extruded aluminium profiles

TF KnowNet

Anodizing is a process in which the thickness of the (natural) oxidation layer on aluminium is electrolytically increased. Aluminium profiles are anodized in large automatic plants, in which the profiles are manually loaded to the process and finally unloaded. Significant costs are related to labor, energy, chemicals and their handling, rent, maintenance, and initial investment. – Based on the process it can be concluded that the costs can be allocated to products based on their area, weight and oxidation layer thickness.

Cost data obtained from plant cost data collection system for various profiles:

	A	B	C	D
1	Data			
2	[µm]	[mm]	[kg]	[euro/m2]
3	Thickne	Perimeter	Weight	Cost
4	5	20	0.09	5.6
5	10	36	0.2	5.2
6	15	65	0.6	5.4
7	20	117	1.6	5.4
8	5	210	3.9	4.1
9	10	378	9.8	4.1
10	15	680	24.4	4.2
11	20	65	0.10	3.8



eit Manufacturing

Aalto University School of Engineering

27

27

Example: cost of anodizing extruded aluminium profiles

TF KnowNet

A view from an automated anodizing plant

Extruded aluminium profiles

eit Manufacturing

Aalto University School of Engineering

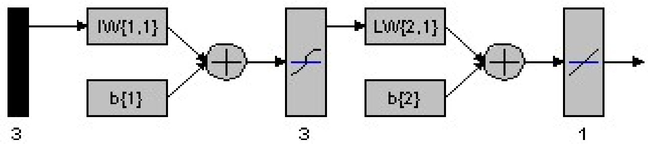
28

28

Example: cost of anodizing extruded aluminium profiles

TF KnowNet

Chosen network topology:



Weights for trained network (Matlab) and a simulation result on spread sheet:

	A	C	E	G	I	K	M	O
1	Input layer	Hidden layer				Output layer		
2	Inputs *p*	Weights *w*	Bias *b*	Weighed sums	Tan-Sigmoid	Weights *w*	Bias *b*	Output *a*
3	5	-5.22	12.2655	-8.79	-1.000	-0.7	4.08	4.926
4		0.31				0.3		=I3*K3+I9*K4+I15*K5+M3
5		-13.2956				0.5		
6								
7								
8								
9	20	0.36	-0.8176	-1.42	-0.890			
10		-0.25			=(1-EXP(-2*G9))/1/(1+EXP(-2*G9))			
11		29.71						
12								
13								
14								
15	0.09	0.92	19.1583	13.01	1.000			
16		-0.61		=C15*A3+C16*A9+C17*A15+E15				
17		15.46						

Aalto University School of Engineering

29

Example: cost of anodizing extruded aluminium profiles

TF KnowNet

Coefficients and R^2 for a linear regression model (Excel) :

	A	B	C
1	SUMMARY OUTPUT		
2			
3	*Regression Statistics*		
4	Multiple R	0.695649449	
5	R Square	0.483928156	
6	Adjusted R Square	0.387164685	
7	Standard Error	0.792960684	
8	Observations	20	
9			
10	ANOVA		
11		*df*	*SS*
12	Regression	3	9.433959736
13	Residual	16	10.06058635
14	Total	19	19.49454608
15			
16		*Coefficients*	*Standard Error*
17	Intercept	4.456101086	0.51808277
18	Thickness	0.048485771	0.032461021
19	Perimeter	-0.007135217	0.002209216
20	Weight	0.187033182	0.069502135
21			
22			
23			
24	RESIDUAL OUTPUT		
25			
26	*Observation*	*Predicted Cost*	*Residuals*
27	1	4.572658579	1.008439342

Example calculation:

0.048*5 – 0.0071*20 + 0.18*0.09 + 4.45 = 4.57

Coefficiens for standardized data:

15		
16		*Coefficients*
17	Intercept	2.31755E-16
18	Thickness	0.27453502
19	Perimeter	-1.368670503
20	Weight	1.130921913
21		

eit Manufacturing

Aalto University School of Engineering

30

Example: cost of anodizing extruded aluminium profiles

TF KnowNet

- For NN training data R^2 = 0.997. For linear regression model R^2 obtained is 0.484.
- In this simple example the number of weights is 16, although we have only three inputs and one output. In practice commercial neural network software can be used (e.g. Matlab) to find suitable network design and values for weights.
- When the network has been trained and values for weights obtained, it is easy to implement. As the parameters of cost calculation or the model itself are not (and should not be) adjusted frequently (typically 1 – 2 times a year), the model could be integrated to cost a calculation program or e.g. to a spread sheet application.
- Many other, e.g. process control applications, can be realized similarly
- Applications requiring more frequent retraining could be for example object recognition or product demand prediction.

eit Manufacturing

Aalto University
School of Engineering

31

31

TF KnowNet

Aalto University
School of Engineering

Thank you

For questions please contact:
Prof. Esko NIEMI (esko.niemi@aalto.fi)

TF KnowNet Consortium:

LMS Laboratory for Manufacturing Systems & Automation

VOLVO VOLVO GROUP

Volkswagen

PRIMA INDUSTRIE

POLITECNICO MILANO 1863

Technische Universität Braunschweig

Aalto University

MONDRAGON UNIBERTSITATEA

eit Manufacturing

32