Regression

Lecture 4 Instruction – without solutions
(Lisa Mannel)

IDS-I-L4

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Linear Regression – Introduction

Height (cm)	Weight (kg)
160	57
162	54
167	64
175	71
175	73
180	70
180	76
182	74
185	82
190	94

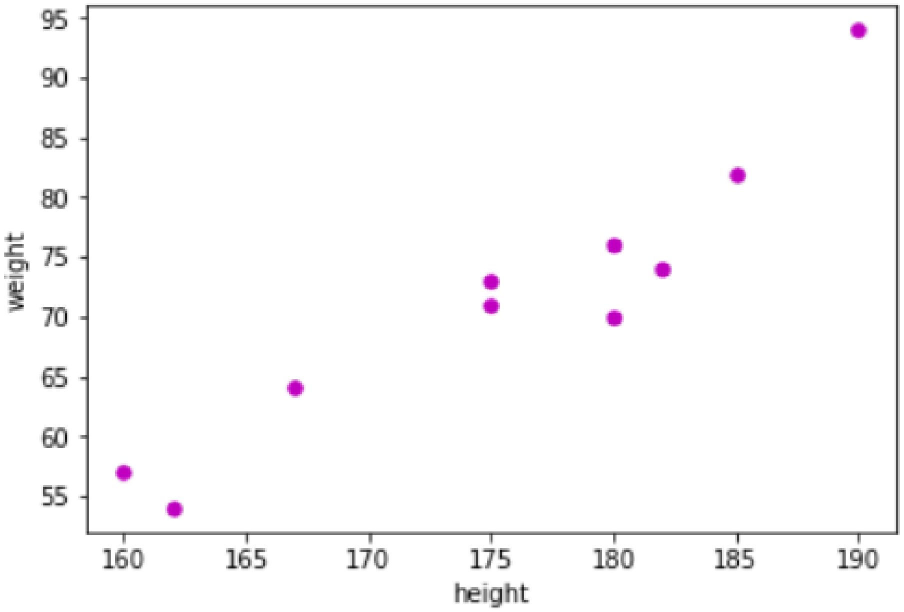
Goal:
Find a linear regression model that predicts the weight of a person (target feature) based on their height (descriptive feature).

Chair of Process and Data Science

Linear Regression – Introduction

Height (cm)	Weight (kg)
160	57
162	54
167	64
175	71
175	73
180	70
180	76
182	74
185	82
190	94

Plotting the data:



Chair of Process and Data Science

Linear Regression – Introduction

Height (cm)	Weight (kg)
160	57
162	54
167	64
175	71
175	73
180	70
180	76
182	74
185	82
190	94

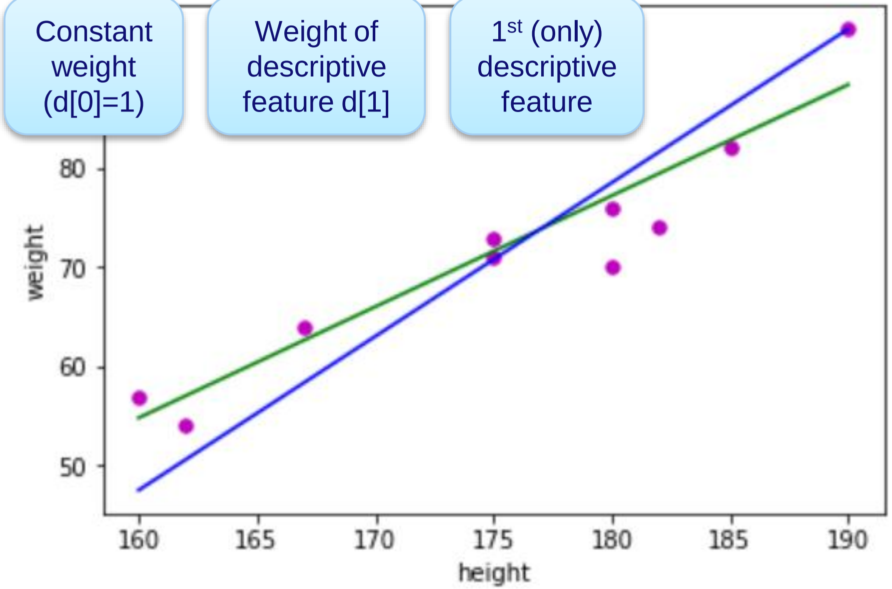
Try to fit a linear function to our data points:

$$\mathbb{M}_{\mathbf{w}}(d) = \mathbf{w}[0] + \mathbf{w}[1] \times \mathbf{d}[1]$$

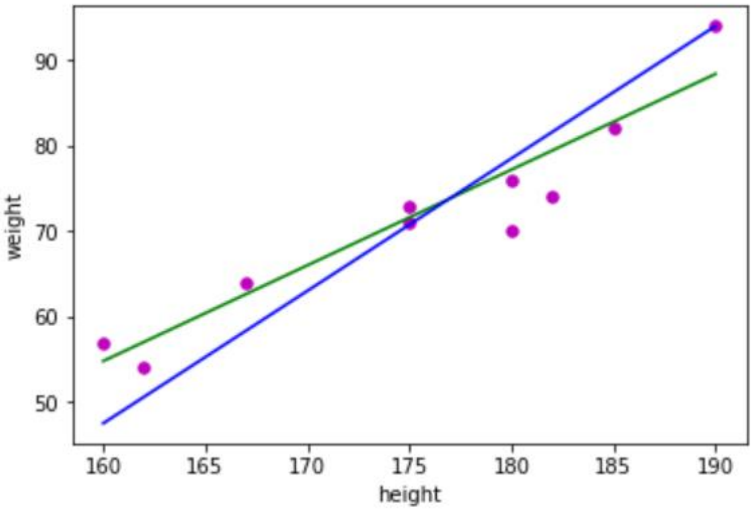
Constant weight (d[0]=1)

Weight of descriptive feature d[1]

1st (only) descriptive feature



Linear Regression – Introduction



But which linear function fits our data best?

$$\mathbb{M}_{\mathbf{w}}(d) = -200.5 + 1.55 \times \mathbf{d}[1]$$
$$\mathbb{M}_{\mathbf{w}}(d) = -124.41 + 1.12 \times \mathbf{d}[1]$$

The function with the smallest error!

Computing the error
Sum of squared errors

$$\mathbb{M}_{\mathbf{w}}(d) = -200.5 + 1.55 \times \mathbf{d}[1]$$
$$\mathbb{M}_{\mathbf{w}}(d) = -124.41 + 1.12 \times \mathbf{d}[1]$$

Height ***d[1]***	**Weight** ***t***
160	57
162	54
167	64
175	71
175	73
180	70
180	76
182	74
185	82
190	94

Exercise 1

Consider the data table on the left and the two linear functions above. The functions predict the weight based on the height.

a) For each function, compute the sum of squared errors.

b) Which function predicts the data best?

Chair of Process and Data Science

Computing the error

Sum of squared errors

$$\mathbb{M}_{\mathbf{w}}(d) = -48 + 0.67 \times \mathbf{d}[1]$$

Height *d[1]*	Weight *t*
160	57
162	54
167	64
175	71
175	73
180	70
180	76
182	74
185	82
190	94

Exercise 2

Consider the data table on the left and the linear functions above. The function predicts the weight based on the height.

Compute the sum of squared errors for this function.

Chair of Process and Data Science

Computing the error

Other error metrics – simple examples in the lecture

$$\text{mean squared error} = \frac{\sum_{i=1}^{n}(t_i - \boxed{\mathbb{M}_{\mathbf{w}}(\mathbf{d}_i)})^2}{n}$$

$$\text{root mean squared error} = \sqrt{\frac{\sum_{i=1}^{n}(t_i - \boxed{\mathbb{M}_{\mathbf{w}}(\mathbf{d}_i)})^2}{n}}$$

$$\text{mean absolute error} = \frac{\sum_{i=1}^{n}\text{abs}(t_i - \boxed{\mathbb{M}_{\mathbf{w}}(\mathbf{d}_i)})}{n}$$

descriptive features

model

Observed value

Predicted value

Chair of Process and Data Science

Computing the error
Multiple descriptive features

Baby Data

Weight [kg]	Size [cm]	Age [months]
2.5	46.4	0
4.2	54.4	0
2.4	45.4	0
3.8	52.9	0
3.2	50.4	1
5.4	59.6	1
3	49.2	1
4.9	56.9	1
4.4	56.7	3
7.4	65.4	3
4.2	55.4	3
6.7	63.4	3
6.2	63.4	6
9.5	72.3	6
5.8	61.8	6
8.7	70.2	6

$$\mathbb{M}_{\mathbf{w}}(d) = -8.4 + 1 \times \text{ WEIGHT } + 0.12 \times \text{ SIZE}$$

Exercise 3

Consider the data table on the left and the linear function above. The function predicts the age of a baby based on the baby's weight and size.

Compute the sum of squared errors for this data and function.

Chair of Process and Data Science

Linear Regression
Gradient Descent

Recall the lecture:

1: **w** $\leftarrow$ random starting point in the weight space
2: **repeat**
3: **for** each **w**[*j*] in **w do**
4: **w**[*j*] $\leftarrow$ **w**[*j*] $+ \alpha \times$ **errorDelta**($\mathcal{D}$, **w**[*j*])
5: **end for**
6: **until** convergence occurs

Randomized weights

$$\mathbb{M}_{\mathbf{w}}(d) = -8.4 + 1 \times \text{ WEIGHT } + 0.12 \times \text{ SIZE}$$

Partial derivatives of the error function *in negative direction*

$$L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) = \frac{1}{2}\sum_{i=1}^{n}(t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i))^2$$

Learning rate (speed, step-size)

$$\alpha = 0.0001$$

Exercise 4

Perform one step of gradient descent (code lines 3 to 5) based on the random weights, learning rate and error function (sum of squared errors) given in gray color. Use the first 3 entries of the **Baby Data**.

a) Give the linear function defined by the updated weights.
b) Compare the error of the input function with the error of the updated function (using sum of squared errors).

Chair of Process and Data Science

Linear Regression
Gradient Descent

(Baby data: 3 entries)

Weight [kg]	Size [cm]	Pred (Age)	Age [months]	Error
3.2	50.4	0.848	1	0.15
4.4	56.7	2.804	3	0.20
6.2	63.4	5.408	6	0.59

$$\mathbb{M}_{\mathbf{w}}(d) = -8.4 + 1 \times \text{ WEIGHT } + 0.12 \times \text{ SIZE}$$

Exercise 5

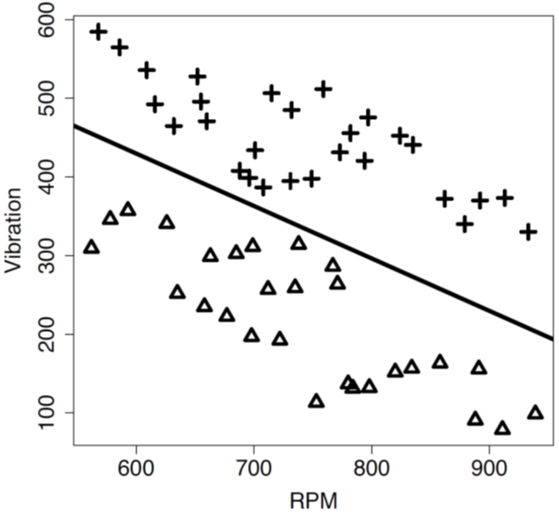
Perform one step of gradient based on the random weights defined by the function above, a learning rate of 0.01, the sum of squared errors and the data given above.

a) Give the linear function defined by the updated weights.

b) Compare the error of the input function with the error of the updated function (using sum of squared errors).

Chair of Process and Data Science

Logistic Regression



For categorical target features: the function is *separating* rather than predicting!

Function shown in the picture:

$$830 - 0.667 \times \text{RPM} - \text{VIBRATION} = 0$$

Logistic function:

- maps the values return by the regression function to the interval $[0, 1]$
- 0 is mapped to 0.5
- values close to 1 are predicted as one class, values close to 0 as the other class
- the closer to 0.5 the more *unsure* is the classification
- usually a *threshold* is used to determine classification (see lecture on evaluation)

Logistic Regression

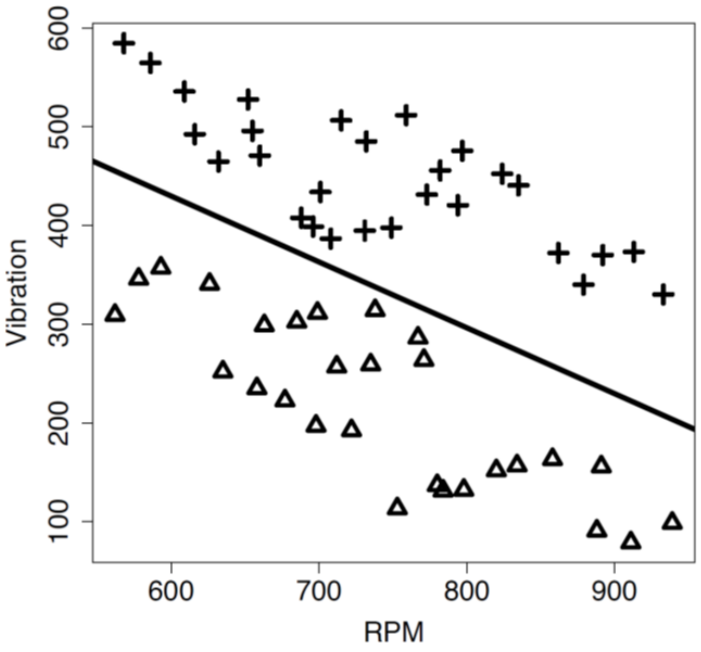
Exercise 6

For RPM $= x_1$ and VIBRATION $= x_2$, consider the logistic regression function

$$\text{Logistic}_1(x_1, x_2) = \frac{1}{1 + e^{-(830-0{,}667\cdot x_1 - x_2)}}.$$

The function classifies the data point $(x_1, x_2) = (800, 400)$ as $+$. How does it classify the data point $(x_1, x_2) = (600, 500)$?

(Assume that all values above 0.5 are classified as one class, and below or euqal as the other class.)



Chair of Process and Data Science

Logistic Regression

	Weight [kg]	Size [cm]	Age [months]	Class	Class prediction
1	2.5	46.4	0	underweight	underweight
2	4.2	54.4	0	overweight	
3	2.4	45.4	0	underweight	
4	3.8	52.9	0	overweight	
5	3.2	50.4	1	underweight	
6	5.4	59.6	1	overweight	
7	3	49.2	1	underweight	
8	4.9	56.9	1	overweight	
9	4.4	56.7	3	underweight	
10	7.4	65.4	3	overweight	
11	4.2	55.4	3	underweight	
12	6.7	63.4	3	overweight	
13	6.2	63.4	6	underweight	
14	9.5	72.3	6	overweight	
15	5.8	61.8	6	underweight	
16	8.7	70.2	6	overweight	

Exercise 7
Consider the data to the left and the below function used for classifcation. Assume a threshold of 0.5 as in the previous exercises.

$$\mathbb{M}_{\mathbf{w}}(d) = \mathrm{Logistic}(-8.4 + 1 \times \mathrm{WEIGHT} + 0.12 \times \mathrm{SIZE} - 1.5 \times \mathrm{AGE})$$

1) Fill in the remaining class predictions.
2) Is the given function perfectly separating the classes?
3) Give an example of a data point, which is classified as underweight (overweight) with high probability.
4) Give an example of a data point, where the classification is unsure.
5) How would you interpret the result of 0.5 with respect to a babys weight?

Chair of Process and Data Science

SVM

Lecture 5 Instruction – without solutions
(Lisa Mannel)

IDS-I-L5

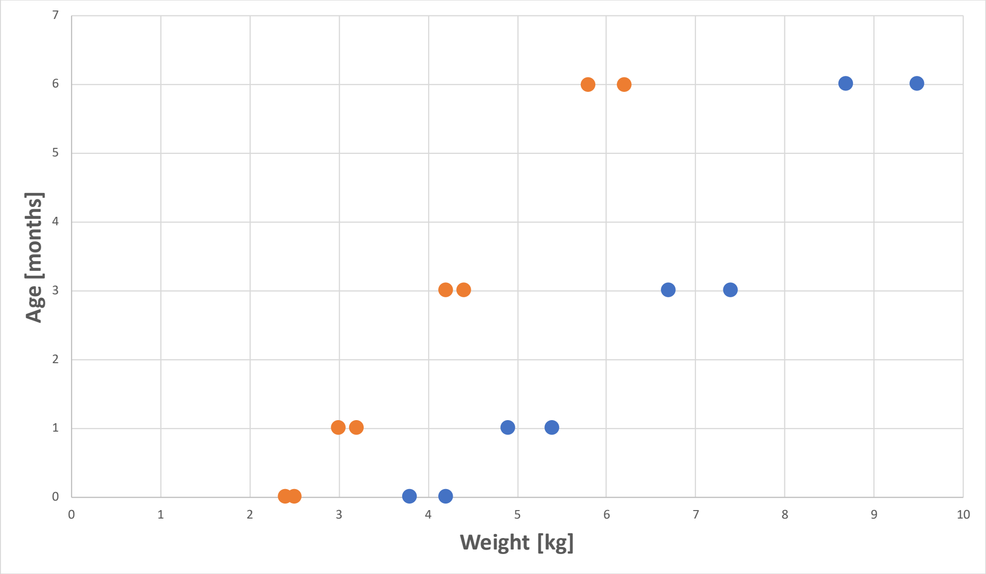
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

SVM – Introduction

Weight [kg]	Age [months]	Class
2.5	0	underweight
4.2	0	overweight
2.4	0	underweight
3.8	0	overweight
3.2	1	underweight
5.4	1	overweight
3	1	underweight
4.9	1	overweight
4.4	3	underweight
7.4	3	overweight
4.2	3	underweight
6.7	3	overweight
6.2	6	underweight
9.5	6	overweight
5.8	6	underweight
8.7	6	overweight

Plotting the data:

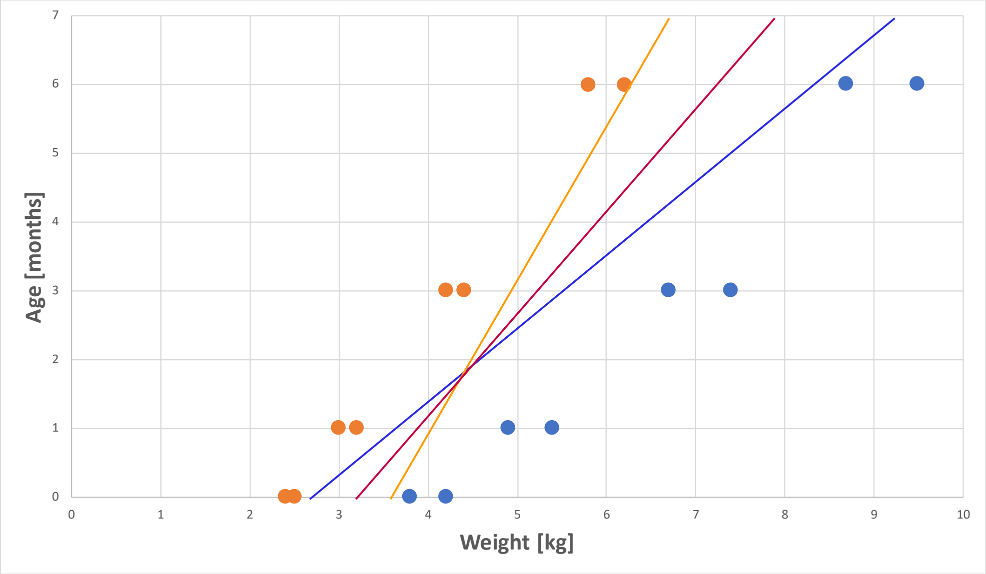


Chair of Process and Data Science

SVM – Introduction

2 descriptive features → hyperplane has dimension 1 (a line)

$$\mathrm{AGE} = 1.3 \cdot \mathrm{WEIGHT} - 4.174$$
$$\mathrm{AGE} = 2.0 \cdot \mathrm{WEIGHT} - 7.6$$
$$\mathrm{AGE} = 0.9 \cdot \mathrm{WEIGHT} - 2.5$$



Chair of Process and Data Science

SVM – Introduction

Different notations are possible, e.g:

$$\text{AGE} = 1.3 \cdot \text{WEIGHT} - 4.174$$
$$\text{AGE} = 2.0 \cdot \text{WEIGHT} - 7.6$$
$$\text{AGE} = 0.9 \cdot \text{WEIGHT} - 2.5$$

$$0 = 1.3 \cdot \text{WEIGHT} - \text{AGE} - 4.174$$
$$0 = 2.0 \cdot \text{WEIGHT} - \text{AGE} - 7.6$$
$$0 = 0.9 \cdot \text{WEIGHT} - \text{AGE} - 2.5$$

$$\vec{w} = (w_1, w_2) = (1.3, -1.0), b = -4.174$$
$$\vec{w} = (w_1, w_2) = (2.0, -1.0), b = -7.6$$
$$\vec{w} = (w_1, w_2) = (0.9, -1.0), b = -2.5$$

Chair of Process and Data Science

SVM – Simple Example

Weight [kg]	Size [cm]	Age [months]	Class
2.5	46.4	0	underweight
4.2	54.4	0	overweight
2.4	45.4	0	underweight
3.8	52.9	0	overweight
3.2	50.4	1	underweight
5.4	59.6	1	overweight
3	49.2	1	underweight
4.9	56.9	1	overweight
4.4	56.7	3	underweight
7.4	65.4	3	overweight
4.2	55.4	3	underweight
6.7	63.4	3	overweight
6.2	63.4	6	underweight
9.5	72.3	6	overweight
5.8	61.8	6	underweight
8.7	70.2	6	overweight

Exercise 9

Given the data table to the left, formulate the optimization problem (see below) for calculating a perfectly separating hyperplane based on the descriptive features *weight* and *age*. Give the constraints such that *overweight* is mapped to positive instances.

$$\min_{\vec{w},b} \frac{1}{2} \|\vec{w}\|^2 \text{ such that}$$

$$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1 \text{ for any } i$$

Chair of Process and Data Science

SVM – Simple Example

Hotel Cost	Distance from Center	Visitor Max Budget	Hotel is booked?
25	5	50	**y**
110	2	30	**y**
49	6	100	**y**
123	11	200	**n**
88	15	300	**n**
41	2	20	**n**
67	10	35	**n**
93	5	40	**n**
29	3	40	**y**
158	1	70	**n**

Exercise 10

Given the data table to the left, formulate the optimization problem (see below) for calculating a perfectly separating hyperplane based on the descriptive features *Hotel Cost, Distance from Center,* and *Visitor Max Budget*. Give the constraints such that *y* is mapped to positive instances.

$\min_{\vec{w},b} \frac{1}{2} \|\vec{w}\|^2$ **such that**

$y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1$ **for any** i