Regression

Lecture 4 Instruction

IDS-I-L4

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Simple Linear Regression Examples

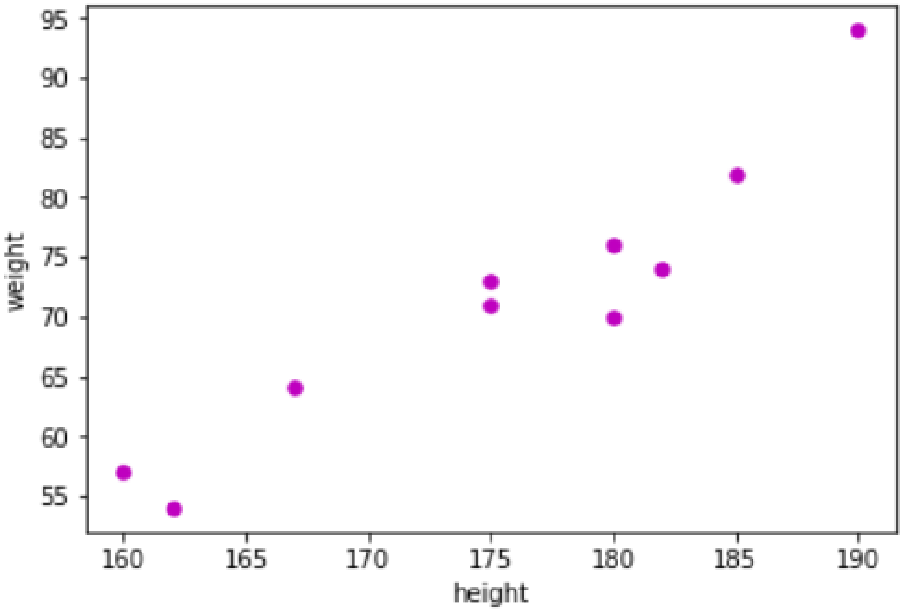
Height (cm)	Weight (kg)
160	57
162	54
167	64
175	71
175	73
180	70
180	76
182	74
185	82
190	94

Exercise:
Calculate a linear regression model that predicts the weight of a person based on their height.

Simple Linear Regression Example

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

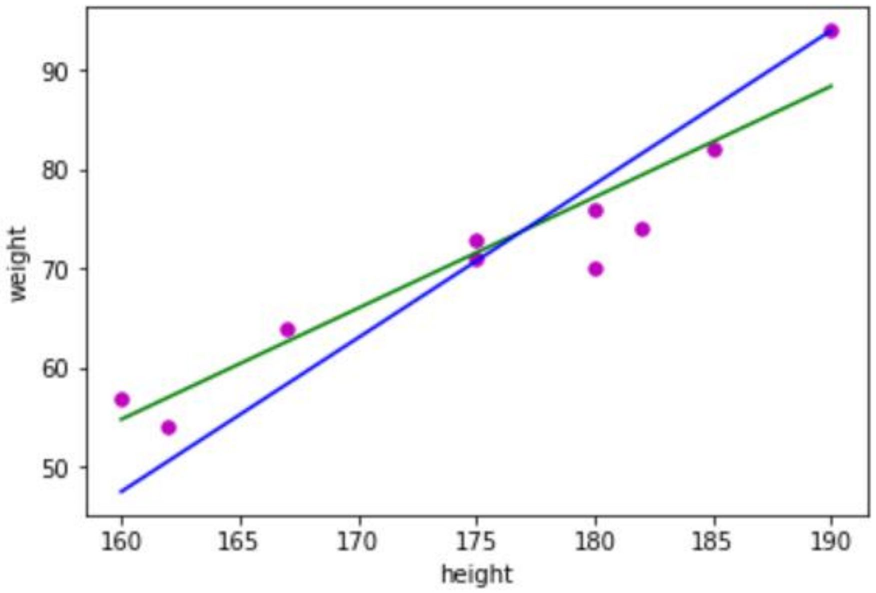
First we plot the data



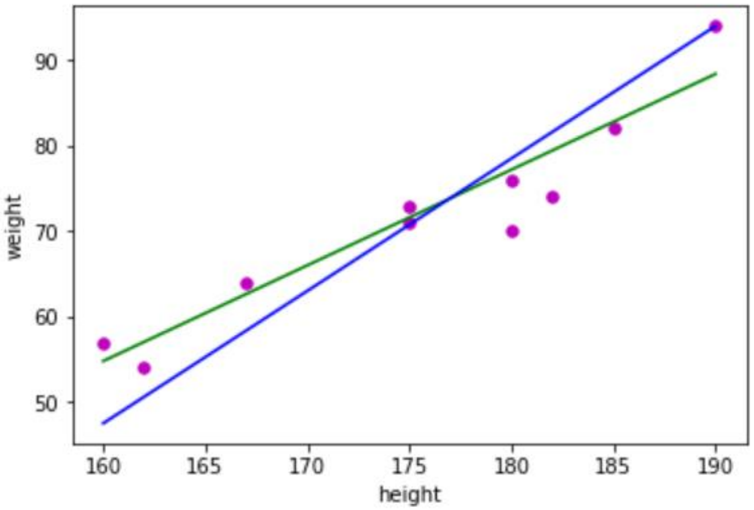
Simple Linear Regression Example

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

We try to fit a linear function to our data points: $y = b + \mathbf{w}\mathbf{x}$



Simple Linear Regression Example



But which line fits our data best?

Blue: $y = -200.5 + 1.55x$

Green: $y = -124.41 + 1.12x$

Basic idea:
Calculate the sum of squared errors and pick the function that reports the smaller value

Simple Linear Regression Example

Blue: $y = -200.5 + 1.55x$

Green: $y = -124.41 + 1.12x$

x	y	Prediction	Error	Error^2	Prediction	Error	Error^2
160	57	47.5	9.5	90.25	54.79	2.21	4.88
162	54	50.6	3.4	11.56	57.03	-3.03	9.18
167	64	58.35	5.65	31.92	62.63	1.37	1.88
175	71	70.75	0.25	0.06	71.59	-0.59	0.35
175	73	70.75	2.25	5.06	71.59	1.41	1.98
180	70	78.5	-8.5	72.25	77.19	-7.19	51.70
180	76	78.5	-2.5	6.25	77.19	-1.19	1.42
182	74	81.60	-7.60	57.76	79.43	-5.43	29.48
185	82	86.25	-4.25	18.06	82.79	-0.79	0.62
190	94	94	0	0	88.39	5.61	31.47

Basic idea:
Calculate the sum of squared errors and pick the function that reports the smaller value

model

data

observed value

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

predicted value

Chair of Process and Data Science

Simple Linear Regression Example

Blue: $y = -200.5 + 1.55x$ $L_2 = 146{,}585$

Green: $y = -124.41 + 1.12x$ $L_2 = 33{,}47$

x	y	Prediction	Error	Error²	Prediction	Error	Error²
160	57	47.5	9.5	90.25	54.79	2.21	4.88
162	54	50.6	3.4	11.56	57.03	-3.03	9.18
167	64	58.35	5.65	31.92	62.63	1.37	1.88
175	71	70.75	0.25	0.06	71.59	-0.59	0.35
175	73	70.75	2.25	5.06	71.59	1.41	1.98
180	70	78.5	-8.5	72.25	77.19	-7.19	51.70
180	76	78.5	-2.5	6.25	77.19	-1.19	1.42
182	74	81.60	-7.60	57.76	79.43	-5.43	29.48
185	82	86.25	-4.25	18.06	82.79	-0.79	0.62
190	94	94	0	0	88.39	5.61	31.47
				293,17			66,94

Basic idea:
Calculate the sum of squared errors and pick the function that reports the smaller value

model

data

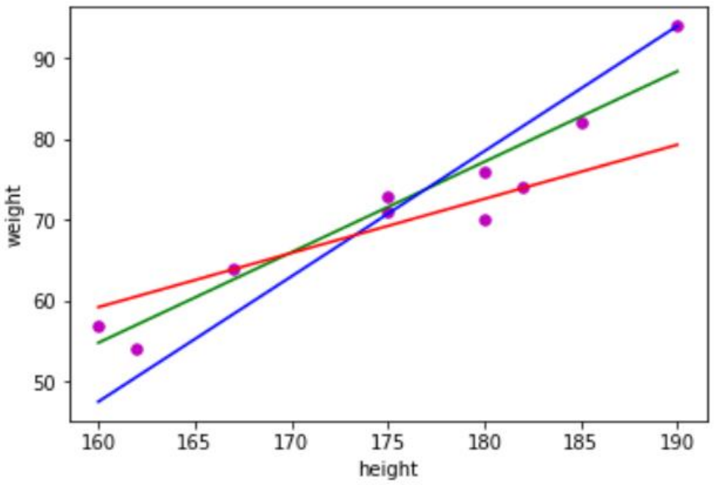
observed value

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

predicted value

Chair of Process and Data Science

Simple Linear Regression Example



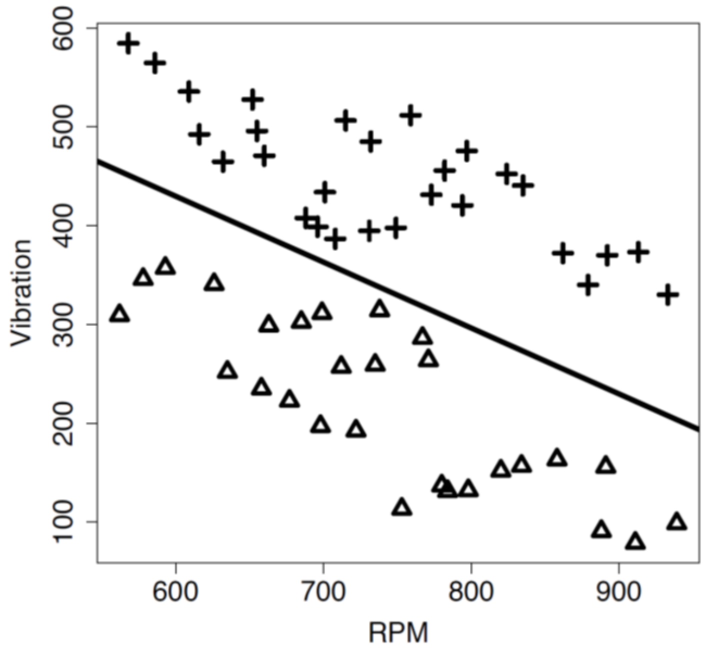
Your turn!

Calculate the sum of squared errors for the red function:

Red: $y = -48 + 0.67x$

Chair of Process and Data Science

Logistic Regression Example



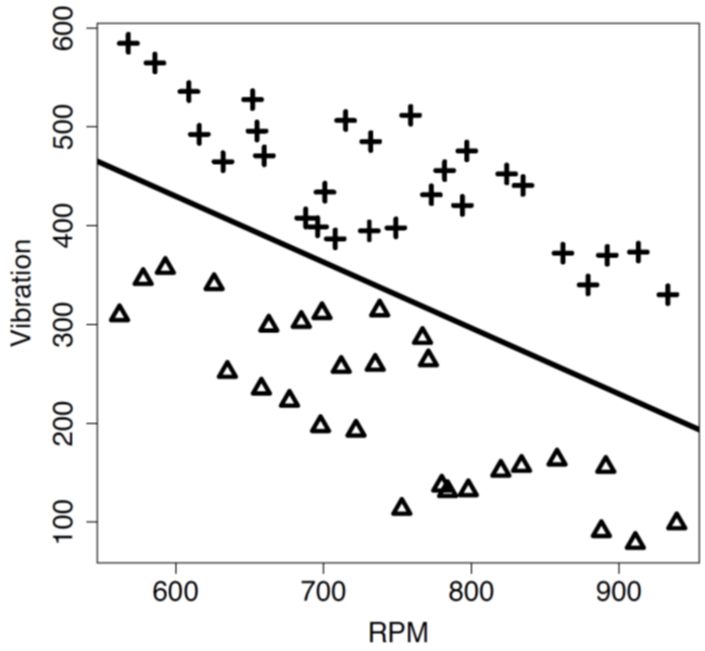
Linear function is *separating* rather than predicting!

Logistic regression function shown in the picture:

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

Exercise:
RPM = 800, Vibration = 400 will be classified as +.
What about RPM = 600, Vibration = 500?

Logistic Regression Example



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

Exercise:
RPM = 800, Vibration = 400 will be classified as +.
What about RPM = 600, Vibration = 500?

Solution:
830 – 0.667 x 800 – 400 = -103.6
→ negative instances are mapped to +

830 -0.667 x 600 – 500 = -70.2
→ mapped to +!

Chair of Process and Data Science

Logistic Regression Example

Your turn:

Exercise:
Consider the logistic regression function $18 + 1.4x - y = 0$ seperating blue from red.
We know that $x = 5$, $y = 15$ is mapped to blue.

What about $x = 8$, $y = 30$?

SVM

Lecture 5 Instruction

IDS-I-L5

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Simple SVM Example

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

Exercise:

Calculate a hyperplane that seperates the class of persons with low credit-worthiness from those with high credit-worthiness based on their monthly income and debts.

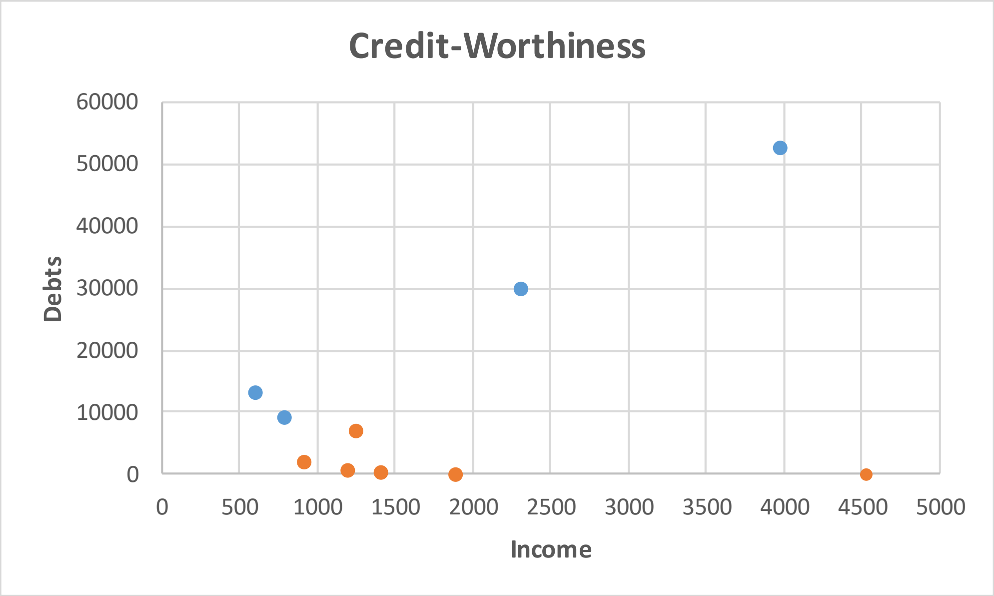
(It is sufficient to formulate the optimization problem ☺)

Chair of Process and Data Science

Simple SVM Example

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

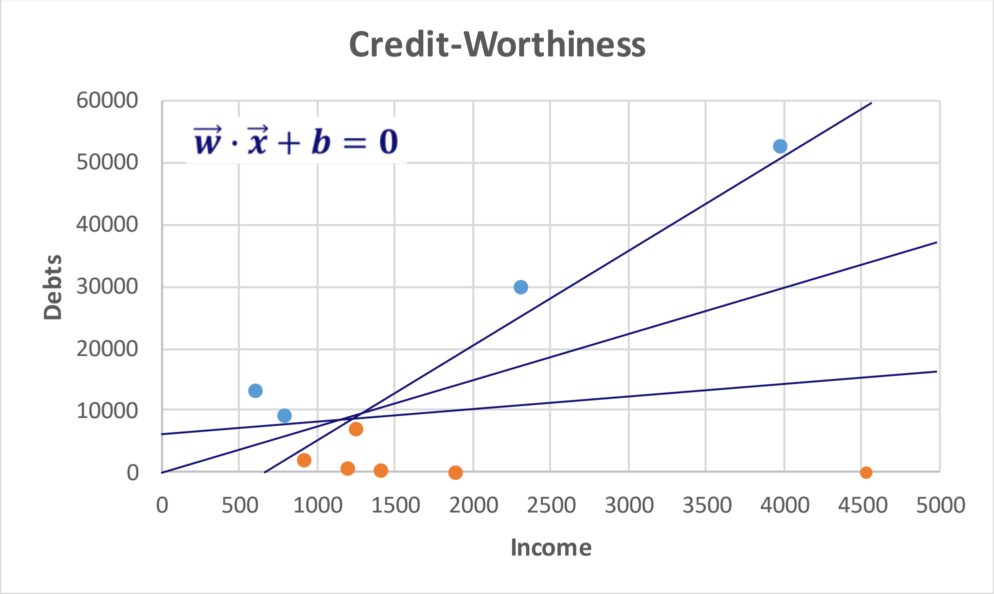
Plot of the data:



Simple SVM Example

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

2 descriptive features
→ hyperplane has dimension 2-1= 1



Chair of Process and Data Science

Simple SVM Example

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

2 descriptive features
→ hyperplane has dimension 2-1= 1

We transform the problem to an optimization problem as shown in the lecture…

Simple SVM Example - Solution

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

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

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

Simple SVM Example - Solution

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

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

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

$$\min \tfrac{1}{2}\|\vec{w}\|^2 = \tfrac{1}{2}(w_1^2 + w_2^2) \text{ such that}$$

$$\begin{aligned}
-1((w_1, w_2) \cdot (1238, 7002) &+ b) \geq 1 \\
-1((w_1, w_2) \cdot (602, 13081) &+ b) \geq 1 \\
-1((w_1, w_2) \cdot (2309, 30002) &+ b) \geq 1 \\
1((w_1, w_2) \cdot (1186, 899) &+ b) \geq 1 \\
-1((w_1, w_2) \cdot (791, 8989) &+ b) \geq 1 \\
1((w_1, w_2) \cdot (1888, 0) &+ b) \geq 1 \\
1((w_1, w_2) \cdot (1400, 421) &+ b) \geq 1 \\
-1((w_1, w_2) \cdot (3971, 52776) &+ b) \geq 1 \\
-1((w_1, w_2) \cdot (910, 2001) &+ b) \geq 1 \\
1((w_1, w_2) \cdot (4522, 10) &+ b) \geq 1
\end{aligned}$$

Simple SVM Example - Solution

Income	Debts	Credit
1238	7002	**Low**
602	13081	**Low**
2309	30002	**Low**
1186	899	**High**
791	8989	**Low**
1888	0	**High**
1400	421	**High**
3971	52776	**Low**
910	2001	**Low**
4522	10	**High**

We don‘t want to solve this manually! ☺

$$\min \tfrac{1}{2}\|\vec{w}\|^2 = \tfrac{1}{2}(w_1^2 + w_2^2) \text{ such that}$$

$$\begin{aligned}
-1((w_1, w_2) \cdot (1238, 7002) &+ b) \geq 1\\
-1((w_1, w_2) \cdot (602, 13081) &+ b) \geq 1\\
-1((w_1, w_2) \cdot (2309, 30002) &+ b) \geq 1\\
1((w_1, w_2) \cdot (1186, 899) &+ b) \geq 1\\
-1((w_1, w_2) \cdot (791, 8989) &+ b) \geq 1\\
1((w_1, w_2) \cdot (1888, 0) &+ b) \geq 1\\
1((w_1, w_2) \cdot (1400, 421) &+ b) \geq 1\\
-1((w_1, w_2) \cdot (3971, 52776) &+ b) \geq 1\\
-1((w_1, w_2) \cdot (910, 2001) &+ b) \geq 1\\
1((w_1, w_2) \cdot (4522, 10) &+ b) \geq 1
\end{aligned}$$

Chair of Process and Data Science

Simple SVM Example

Hotel Cost	Distance from Center	Visitor Max Budget	Hotel 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**

Your turn!

Formulate the optimization problem for finding the best hyperplane classifying the given data!