



Funded by
the European Union



Course on Artificial Intelligence

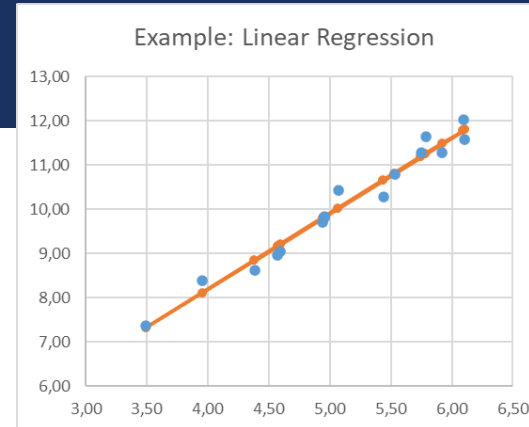
José Coelho, 2024

Lesson 3: Introduction to Machine Learning: Learning from Examples – Part B

Table of contents

1. Linear regression
2. Gradient descent
3. Linear classification
4. K nearest neighbors
5. Combine learning
6. Neural networks

Linear regression



- Loss $L(y, \hat{y})$
 - 0/1: $L_{0/1}(y, \hat{y}) = 0$ se $y = \hat{y}$, 1 c. c.
 - Absoluta: $L_1(y, \hat{y}) = |y - \hat{y}|$
 - Quadrática: $L_2(y, \hat{y}) = (y - \hat{y})^2$

■ Univariate linear regression

- $h_w(x) = w_1x + w_0$
- $w_1 = \frac{N(\sum x_j y_j) - (\sum x_j)(\sum y_j)}{N(\sum x_j^2) - (\sum x_j)^2}$
- $w_0 = (\sum y_j - w_1(\sum x_j)) / N$

w1	1,718
w0	1,308

Case	x	y	x ²	xy	h(x)	L2	L1
1	5,07	10,42	25,71	52,85	10,02	0,16	0,40
2	4,96	9,84	24,56	48,77	9,82	0,00	0,02
3	5,78	11,63	33,46	67,29	11,25	0,15	0,38
4	3,96	8,39	15,65	33,20	8,11	0,08	0,29
5	5,44	10,29	29,59	55,95	10,66	0,14	0,37
6	4,38	8,62	19,23	37,81	8,84	0,05	0,22
7	4,94	9,70	24,41	47,93	9,80	0,01	0,10
8	5,53	10,79	30,58	59,65	10,81	0,00	0,02
9	6,11	11,57	37,28	70,66	11,80	0,05	0,23
10	3,50	7,36	12,22	25,73	7,31	0,00	0,05
11	5,92	11,29	35,07	66,85	11,48	0,04	0,19
12	6,09	12,04	37,13	73,34	11,78	0,07	0,26
13	4,57	8,97	20,92	41,02	9,17	0,04	0,20
14	4,96	9,81	24,59	48,66	9,83	0,00	0,02
15	4,60	9,05	21,12	41,61	9,20	0,02	0,15
16	5,75	11,29	33,05	64,89	11,19	0,01	0,10
Sum	81,56	161,06	424,57	836,20	Average	0,05	0,19

$$f(x) = x * 1,5 + 2 + Uniform(0; 1)$$

Gradient descent

- Initialize w_0, w_1
- Until it converges
 - $w_0 \leftarrow w_0 + \alpha \sum (y_j - h_w(x_j))$
 - $w_1 \leftarrow w_1 + \alpha \sum (y_j - h_w(x_j)) \times x_j$

	y-h(x)	(y-h(x))x	y-h(x)	(y-h(x))x	y-h(x)	(y-h(x))x	y-h(x)	(y-h(x))x
	10,42	52,85	1,62	8,22	0,57	2,89	0,44	2,25
	9,84	48,77	1,23	6,10	0,20	1,01	0,08	0,40
	11,63	67,29	1,64	9,46	0,44	2,55	0,30	1,73
	8,39	33,20	1,45	5,75	0,62	2,47	0,53	2,08
	10,29	55,95	0,87	4,71	-0,26	-1,42	-0,39	-2,15
	8,62	37,81	0,97	4,24	0,05	0,23	-0,06	-0,25
	9,70	47,93	1,11	5,51	0,09	0,44	-0,03	-0,17
	10,79	59,65	1,21	6,72	0,07	0,39	-0,07	-0,36
	11,57	70,66	1,04	6,35	-0,22	-1,33	-0,37	-2,25
	7,36	25,73	1,19	4,17	0,46	1,60	0,37	1,29
	11,29	66,85	1,06	6,29	-0,16	-0,94	-0,30	-1,81
	12,04	73,34	1,52	9,28	0,27	1,63	0,12	0,71
	8,97	41,02	1,00	4,56	0,04	0,20	-0,07	-0,32
	9,81	48,66	1,20	5,93	0,17	0,83	0,04	0,22
	9,05	41,61	1,05	4,80	0,09	0,41	-0,03	-0,12
	11,29	64,89	1,35	7,77	0,16	0,94	0,02	0,13
	161,06	836,20	19,51	99,87	2,60	11,88	0,58	1,37
w1	0	1,672		1,872		1,896		1,899
w0	0	0,322		0,361		0,366		0,368
alpha	0,002							

Linear classification

■ Cutoff

- $h_w(x) = 1$ se $w_2x_2 + w_1x_1 + w_0 \geq 0$, 0 c.c.

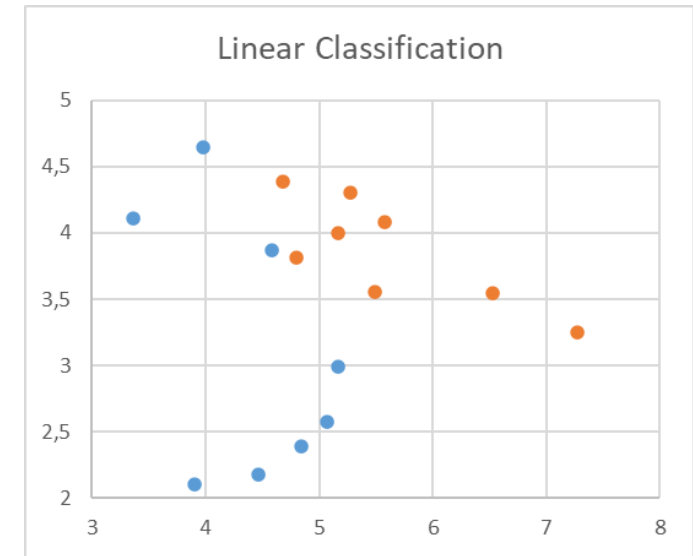
- $w_i \leftarrow w_i + \alpha(y - h_w(x))x_i$

■ Logistics function

- $h_w(x) = \frac{1}{1 + e^{-(w_2x_2 + w_1x_1 + w_0)}}$

- $w_i \leftarrow w_i + \alpha(y - h_w(x))h_w(x)(1 - h_w(x))x_i$

Case	x1	x2	y
1	4,46	2,18	0
2	5,16	2,99	0
3	5,58	4,08	1
4	3,36	4,11	0
5	4,67	4,38	1
6	4,58	3,87	0
7	5,28	4,30	1
8	5,17	4,00	1
9	6,53	3,54	1
10	4,85	2,39	0
11	3,98	4,65	0
12	4,80	3,81	1
13	5,49	3,56	1
14	7,27	3,25	1
15	3,90	2,11	0
16	5,07	2,58	0



w0	w1	w2	h(x)	L2	w0	w1	w2	h(x)	L2	w0	w1	w2	h(x)	L2	w0	w1	w2	h(x)	L2	
0,000	0,000	0,000	0,50	0,25	-0,099	-0,189	-0,062	0,25	0,06	-0,160	-0,214	-0,039	0,23	0,05	-0,224	-0,257	-0,025	0,19	0,04	
-0,063	-0,279	-0,136	0,13	0,02	-0,123	-0,296	-0,115	0,12	0,01	-0,180	-0,306	-0,084	0,12	0,01	-0,240	-0,324	-0,058	0,11	0,01	
-0,070	-0,316	-0,158	0,08	0,85	-0,129	-0,329	-0,133	0,08	0,86	-0,187	-0,338	-0,102	0,08	0,85	-0,245	-0,352	-0,074	0,07	0,86	
-0,037	-0,133	-0,024	0,36	0,13	-0,097	-0,150	-0,002	0,35	0,12	-0,154	-0,155	0,031	0,37	0,13	-0,213	-0,173	0,056	0,36	0,13	
-0,078	-0,272	-0,193	0,10	0,81	-0,137	-0,285	-0,167	0,10	0,81	-0,196	-0,298	-0,143	0,10	0,81	-0,255	-0,314	-0,116	0,10	0,82	
-0,038	-0,082	-0,015	0,38	0,15	-0,097	-0,096	0,010	0,38	0,14	-0,157	-0,112	0,031	0,37	0,13	-0,215	-0,129	0,058	0,36	0,13	
-0,083	-0,290	-0,191	0,08	0,85	-0,141	-0,299	-0,162	0,08	0,84	-0,199	-0,307	-0,133	0,08	0,84	-0,257	-0,318	-0,102	0,09	0,84	
-0,049	-0,111	-0,045	0,31	0,48	-0,107	-0,118	-0,014	0,32	0,47	-0,164	-0,121	0,018	0,33	0,45	-0,221	-0,129	0,052	0,34	0,44	
0,025	0,271	0,250	0,94	0,00	-0,033	0,265	0,282	0,94	0,00	-0,090	0,262	0,314	0,94	0,00	-0,147	0,254	0,348	0,94	0,00	
0,027	0,284	0,257	0,88	0,78	-0,031	0,277	0,288	0,88	0,78	-0,088	0,273	0,321	0,88	0,78	-0,145	0,265	0,354	0,88	0,77	
-0,019	0,062	0,148	0,71	0,51	-0,077	0,053	0,178	0,72	0,52	-0,134	0,049	0,210	0,74	0,55	-0,192	0,039	0,243	0,75	0,56	
-0,092	-0,228	-0,192	0,13	0,76	-0,149	-0,235	-0,159	0,13	0,75	-0,206	-0,234	-0,121	0,14	0,73	-0,262	-0,241	-0,084	0,15	0,72	
-0,043	0,005	-0,006	0,49	0,26	-0,100	0,004	0,031	0,51	0,24	-0,153	0,018	0,079	0,56	0,20	-0,208	0,018	0,121	0,58	0,18	
0,020	0,355	0,221	0,96	0,00	-0,038	0,342	0,250	0,96	0,00	-0,098	0,319	0,274	0,96	0,00	-0,157	0,299	0,304	0,95	0,00	
0,021	0,359	0,222	0,87	0,75	-0,038	0,346	0,252	0,86	0,75	-0,097	0,325	0,277	0,85	0,73	-0,156	0,307	0,307	0,84	0,71	
-0,029	0,166	0,118	0,75	0,57	-0,088	0,148	0,145	0,74	0,54	-0,151	0,116	0,164	0,70	0,49	-0,211	0,090	0,190	0,68	0,46	
alpha:	0,5			0,447					0,432					0,423					0,417	

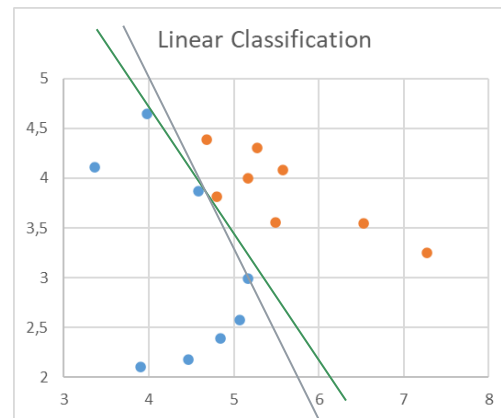
w0	w1	w2	h(x)	L0/1	w0	w1	w2	h(x)	L0/1	w0	w1	w2	h(x)	L0/1	w0	w1	w2	h(x)	L0/1	
0,000	0,000	0,000	1	1	-0,010	-0,008	0,019	0	0	-0,020	-0,023	0,031	0	0	-0,030	-0,034	0,028	0	0	
-0,010	-0,045	-0,022	0	0	-0,010	-0,008	0,019	1	1	-0,020	-0,023	0,031	0	0	-0,030	-0,034	0,028	0	0	
-0,010	-0,045	-0,022	0	1	-0,020	-0,060	-0,011	0	1	-0,020	-0,023	0,031	0	1	-0,030	-0,034	0,028	0	1	
0,000	0,011	0,019	1	1	-0,010	-0,004	0,030	1	1	-0,010	0,032	0,071	1	1	-0,020	0,022	0,069	1	1	
-0,010	-0,022	-0,022	0	1	-0,020	-0,038	-0,011	0	1	-0,020	-0,001	0,030	1	0	-0,030	-0,012	0,028	1	0	
0,000	0,024	0,022	1	1	-0,010	0,009	0,033	1	1	-0,020	-0,001	0,030	1	1	-0,030	-0,012	0,028	1	1	
-0,010	-0,021	-0,017	0	1	-0,020	-0,037	-0,006	0	1	-0,030	-0,047	-0,008	0	1	-0,040	-0,057	-0,011	0	1	
0,000	0,031	0,026	1	0	-0,010	0,016	0,037	1	0	-0,020	0,006	0,035	1	0	-0,030	-0,005	0,032	1	0	
0,000	0,031	0,026	1	0	-0,010	0,016	0,037	1	0	-0,020	0,006	0,035	1	0	-0,030	-0,005	0,032	1	0	
0,000	0,031	0,026	1	1	-0,010	0,016	0,037	1	1	-0,020	0,006	0,035	1	1	-0,030	-0,005	0,032	1	1	
-0,010	-0,017	0,002	0	0	-0,020	-0,032	0,013	0	0	-0,030	-0,043	0,011	0	0	-0,040	-0,053	0,008	0	0	
-0,010	-0,017	0,002	0	1	-0,020	-0,032	0,013	0	1	-0,030	-0,043	0,011	0	1	-0,040	-0,053	0,008	0	1	
0,000	0,031	0,040	1	0	-0,010	0,016	0,052	1	0	-0,020	0,005	0,049	1	0	-0,030	-0,005	0,046	1	0	
0,000	0,031	0,040	1	0	-0,010	0,016	0,052	1	0	-0,020	0,005	0,049	1	0	-0,030	-0,005	0,046	1	0	
0,000	0,031	0,040	1	1	-0,010	0,016	0,052	1	1	-0,020	0,005	0,049	1	1	-0,030	-0,005	0,046	1	1	
-0,010	-0,008	0,019	0	0	-0,020	-0,023	0,031	0	0	-0,030	-0,034	0,028	0	0	-0,040	-0,044	0,025	0	0	
alpha:	0,01			56%					56%					44%					44%	

K nearest neighbors

- Non-parametric models
 - Use past cases
 - Use the K closest examples
 - Use of trees
- Nonparametric regression
 - K nearest neighbors
 - Locally weighted regression
- Support Vector Machines
 - Kernel Functions

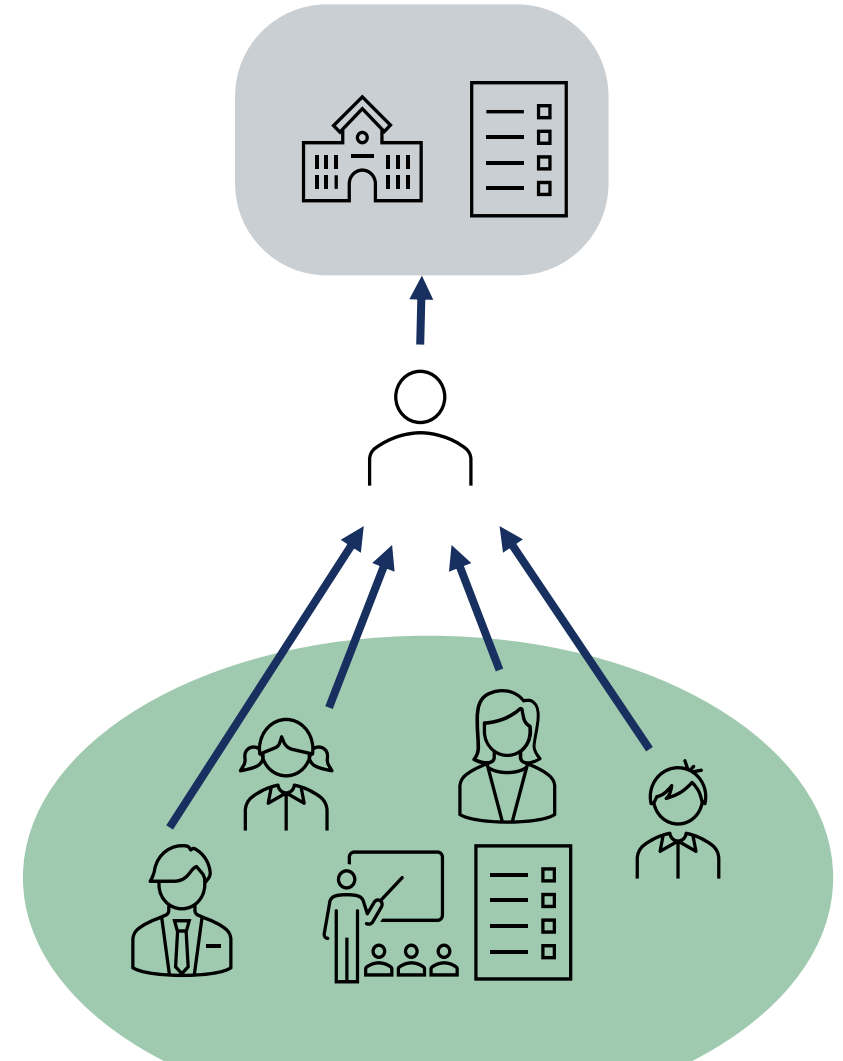
Case	Materials	Formative Activities	Interventions	Evaluations	Grade
1	1	2	2	1	1
2	3	2	1	3	2
3	2	2	1	2	1
4	2	2	1	3	2
5	2	2	2	1	1
6	2	3	1	2	2
7	1	2	3	2	1
8	2	2	3	2	2
9	2	1	1	2	1
10	3	2	3	3	2
11	1	2	2	1	1
12	3	1	2	3	2
13	2	2	1	1	1
14	2	2	2	3	2
15	1	1	3	1	1
16	3	3	3	2	2

Train						
1	0	5	2	3	2	5
2	5	2	3	2	7	4
3	3	4	1	2	5	4
4	4	3	2	1	6	5
5	1	4	1	2	3	4
6	4	5	2	3	6	3
7	2	5	4	3	2	3
8	3	4	3	2	3	2
9	4	3	2	3	4	5
10	5	2	5	2	5	2
Test						
11	12	13	14	15	16	
1º	1	2	1	2	1	2
2º	1	2	1	2	1	2
3º	1	2	1	1	1	2
h(x)	1	2	1	2	1	2
y	1	2	1	2	1	2
L0/1	0	0	0	0	0	0



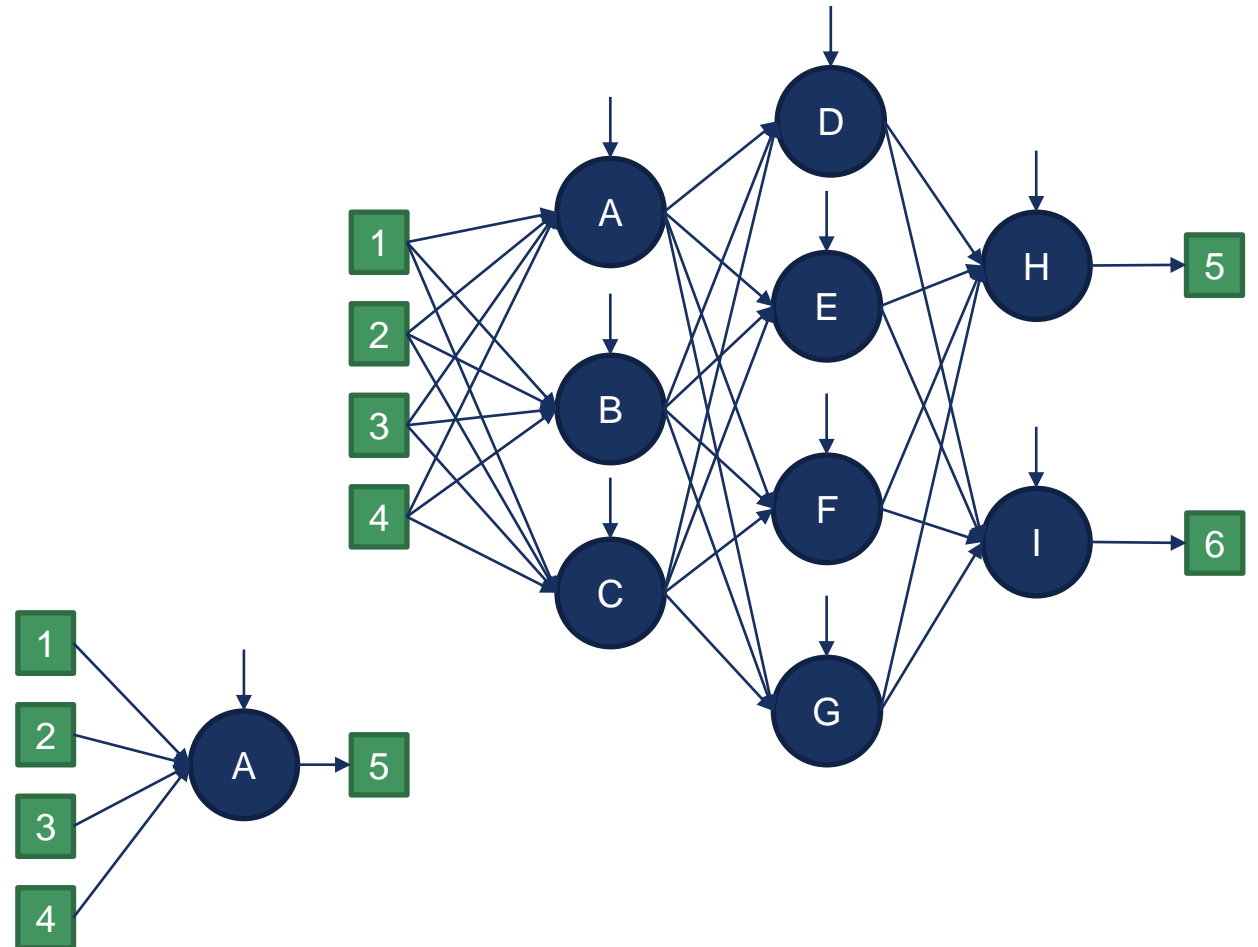
Combine learning

- Bagging
 - Get $h(x)$ based on a random training set
 - Repeat K times, obtaining K functions $h(x)$
 - Aggregate (most frequent or average)
- Random florest
 - Generate K decision trees, with randomness:
 - Select a subset of attributes \sqrt{n} or $\frac{n}{3}$
 - Just some cutoff points analyzed
- Stacking
 - Get K good functions $h(x)$ and add the value to x , leaving x'
 - Get $h(x')$, that use the result of the K functions
- Boosting
 - Whenever a case fails, clone the case



Neural networks

- Deep learning
- Neuron
 - Linear classification
 - Activation function
- Acyclic networks (NN)
 - Entry level
 - Hidden levels
 - Output level
- Train
 - Back-propagation
- Convulsive networks (CNN)
 - Process images!
 - Neighborhood in data (pixels)
 - Spatial invariance!
- Recurring networks (RNN)
 - Process sequential data! (audio, text)
 - Memory



Thank you for attention

Resources:

- Microsoft Power Point / Clipchamp / DeepL
- Russell, S. J. & Norvig, P. (2010). Artificial intelligence: A modern approach (3rd ed). Prentice Hall.



Funded by
the European Union

ASSISTANT - CHALLENGE BASED LEARNING IN AI ENHANCED DIGITAL TRANSFORMATION CURRICULAR NO. 2022-I-LT01-KA220-HED-000086555

