Used Car Pricing via Linear Regression

Presented by: Elijah Nichols
Mentor: Dr. Ovidiu Calin

Abstract

- Current used car pricing method inefficient/flawed
  - Kelley Blue Book
- Humans approximating neural network
- Create simple neural network to approximate market price
    - Feedforward
    - Regression
    - Supervised training

Process

- Find dataset of used cars
  - Kaggle
- Clean dataset
  - Eliminating bad entries, converting existing entries to vectorizable formats
- Create linear regression model
- Expand to neural network
  - Add hidden layer, activation function
- Play with number of neurons and layers

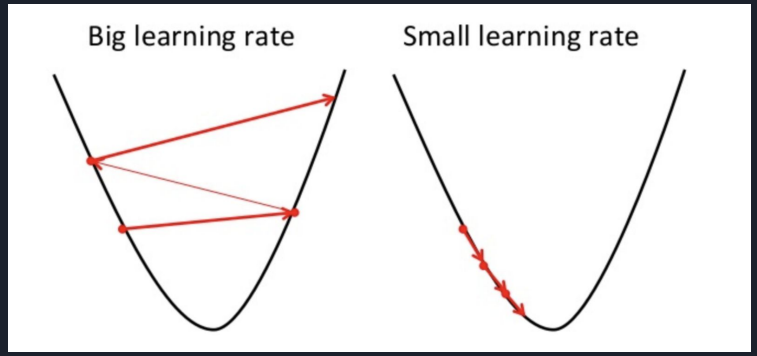
Data

- ~280,000 records after cleaning
- Price, vehicleType, year, manual/auto, model, mileage, fuelType, brand, damage
- Not pipelined
  - Data not fed in from .csv; entire dataset stored in program
- 50-dimensional
  - 9 weights + 1 bias per neuron; 5 neurons

	A	B	C	D	E	F	G	H	I	J	K
1	price	vehicleTyp	yearOfReg	gearbox	powerPS	model	kilometer	fuelType	brand	notRepairedDamage	
2	1650	coupe	2018	1	140	3er	150000	1	bmw	-1	
3	1900	kleinwager	2018	1	75	polo	150000	1	volkswage	-1	
4	11950	bus	2017	1	130	transporte	150000	-1	volkswage	-1	
5	1300	coupe	2017	1	101	3er	150000	1	bmw	0	
6	6000	kleinwager	2017	1	155	punto	150000	1	fiat	-1	
7	800	kleinwager	2017	1	16	a3	30000	1	audi	1	
8	9999	kombi	2017	1	102	a3	90000	1	audi	-1	
9	3150	kombi	2017	1	135	golf	20000	1	volkswage	-1	
10	2350	limousine	2017	-1	170	5er	150000	1	bmw	-1	
11	2350	limousine	2017	-1	170	5er	150000	1	bmw	-1	
12	64280	bus	2016	-1	204	transporte	5000	-1	volkswage	-1	
13	62500	bus	2016	-1	204	transporte	5000	-1	volkswage	-1	
14	59500	bus	2016	-1	150	transporte	5000	-1	volkswage	-1	
15	52500	bus	2016	-1	150	transporte	5000	-1	volkswage	-1	
16	49900	bus	2016	-1	204	transporte	5000	-1	volkswage	-1	
17	49500	bus	2016	-1	190	vito	5000	-1	mercedes_	-1	
18	37500	bus	2016	-1	150	andere	5000	-1	bmw	-1	
19	35590	bus	2016	1	150	transporte	5000	-1	volkswage	-1	
20	34999	bus	2016	1	150	alhambra	5000	-1	seat	-1	
21	34999	bus	2016	1	150	transporte	5000	-1	volkswage	-1	
22	32900	bus	2016	1	150	andere	5000	-1	bmw	-1	
23	32500	bus	2016	1	145	vivaro	5000	-1	opel	-1	

	A	B	C	D	E	F	G	H	I	J
1	vehicleTyp	yearOfReg	gearbox	powerPS	model	kilometers	fuelType	brand	notRepaire	price
2	5	81	1	64	60	30	-1	2	-1	100
3	6	86	1	90	60	1	1	2	-1	100
4	1	80	1	54	60	30	1	2	-1	100
5	1	83	1	60	60	30	1	2	-1	100
6	1	83	1	60	60	30	1	2	-1	100
7	1	85	1	60	60	30	1	2	-1	100
8	1	87	1	60	60	30	1	2	-1	100
9	1	84	1	75	60	30	1	2	-1	100
10	1	87	1	75	60	30	1	2	-1	100
11	1	87	1	0	117	30	1	2	-1	100
12	1	88	1	0	117	30	1	2	-1	100
13	1	87	1	50	117	30	1	2	-1	100
14	1	87	1	50	117	30	1	2	-1	100
15	1	89	1	50	117	30	1	2	-1	100
16	6	84	1	54	117	30	1	2	-1	100
17	2	84	1	55	117	30	1	2	-1	100
18	1	86	1	60	117	30	1	2	-1	100
19	1	86	1	60	117	30	1	2	-1	100
20	1	87	1	60	117	30	1	2	-1	100
21	1	88	1	60	117	30	1	2	-1	100
22	1	91	1	75	215	30	1	2	-1	100
23	6	85	1	75	248	30	1	9	-1	100

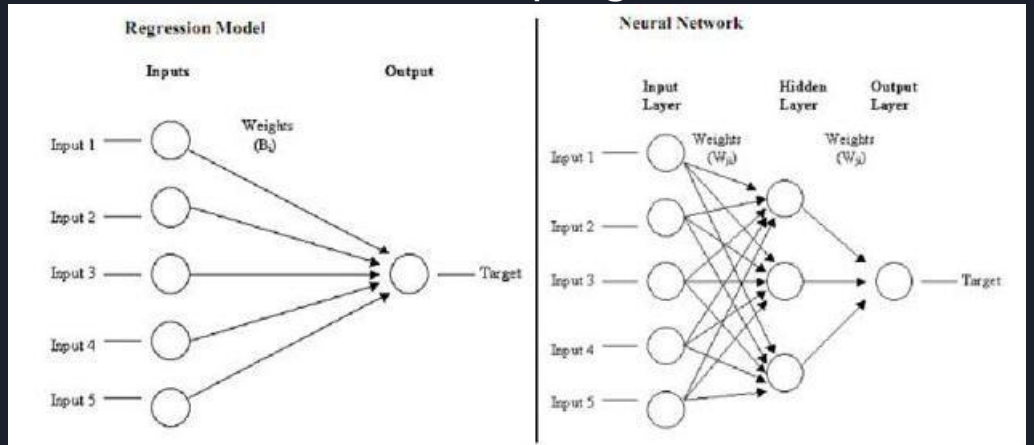
Training



- Cost function
  - output^2-expectedValue^2
- Activation function
  - sigmoid
- Gradient descent
  - Tunes weights and biases to minimize cost function
  - Stochastic stepping toward lowest point in N-dimensional space
  - Taking derivative of equation with dozens, hundreds or more variables; difficult at best

Final Product - A Proper Neural Network

- 1 hidden layer w/ 4 neurons
- Not pipelined; data stored in file
  - Slow
- High error - must teach it better
- Demonstration model for presentation
  - 4,000 entries stored in program



What Was Learned

- Machine learning is a high-level pursuit
  - Don’t get bogged down with code
  - High-level API>low-level
- Crash course in Python
- Data is important
- Basics attributes of neural networks