Mašinsko učenje 1

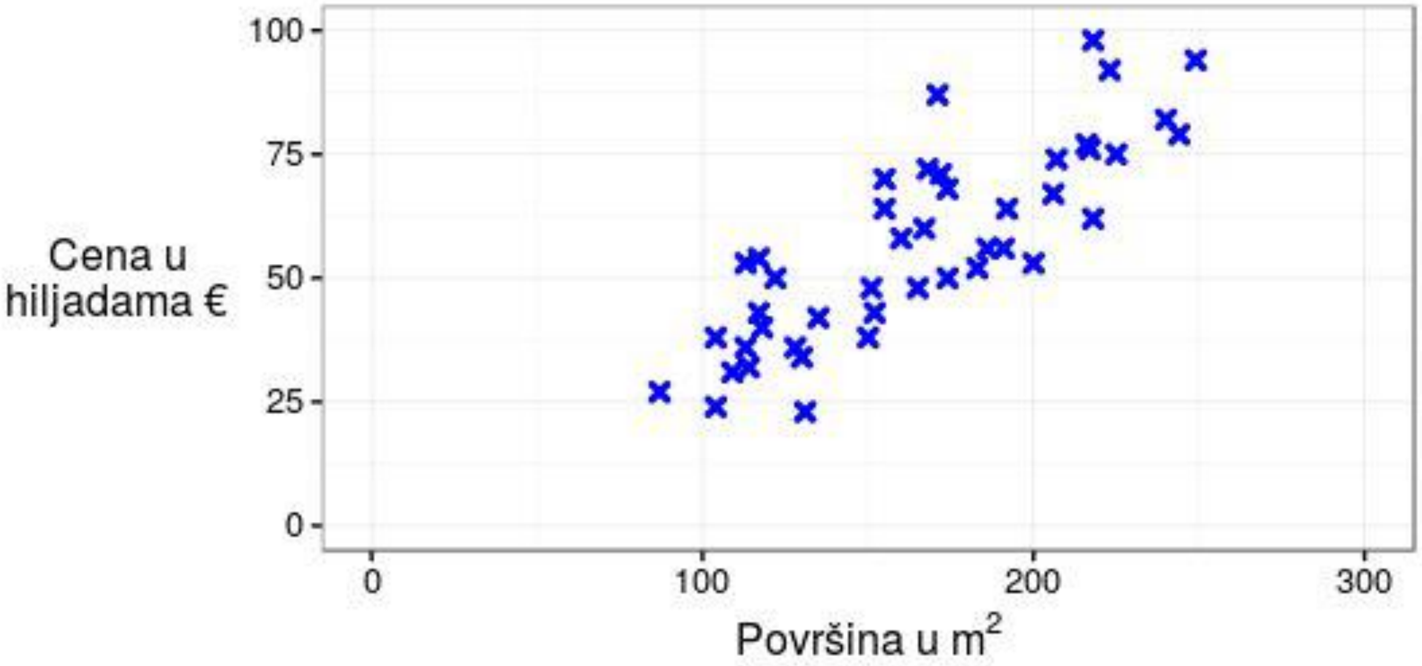
MAS Informatike – Nauka o podacima

Linearna regresija - primer

Predvideti **cenu** stana na osnovu njegove kvadrature.

Trening skup

Površina (u m^2)	Cena (u hiljadama €)
152	43
244	79
104	24
87	27
…	…



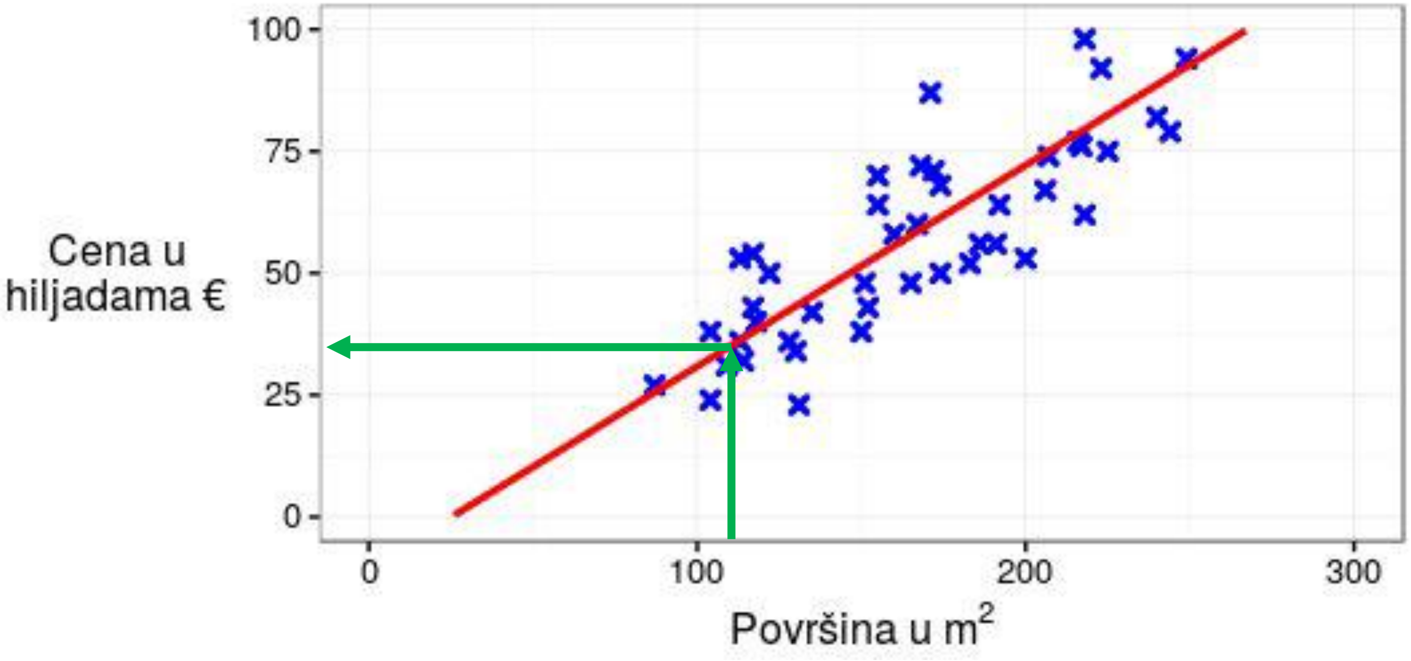
Linearna regresija - primer

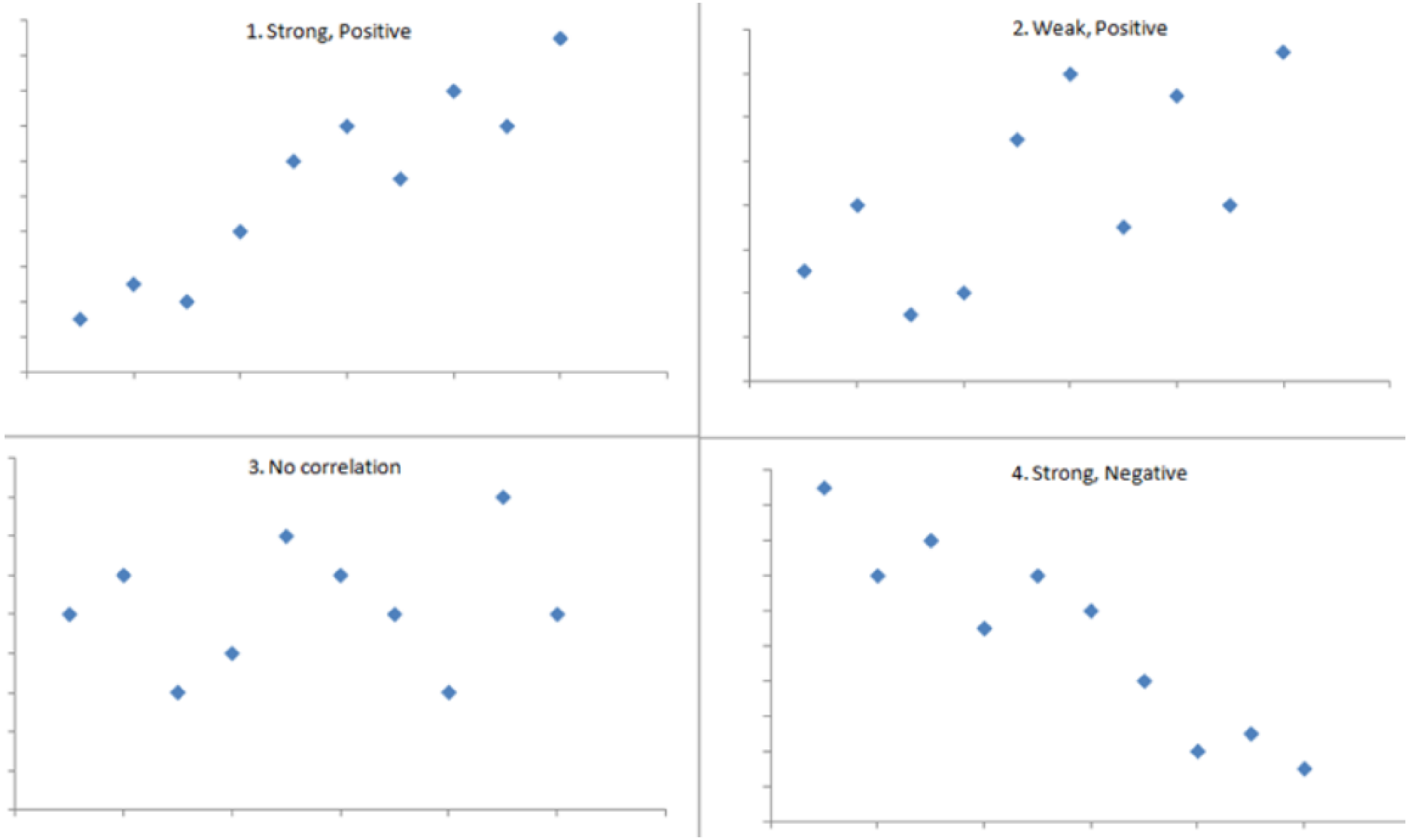
Predvideti **cenu** stana na osnovu njegove kvadrature.

Predviđanje kontinualne vrednosti

Trening skup

Površina (u m^2)	Cena (u hiljadama €)
152	43
244	79
104	24
87	27
…	…



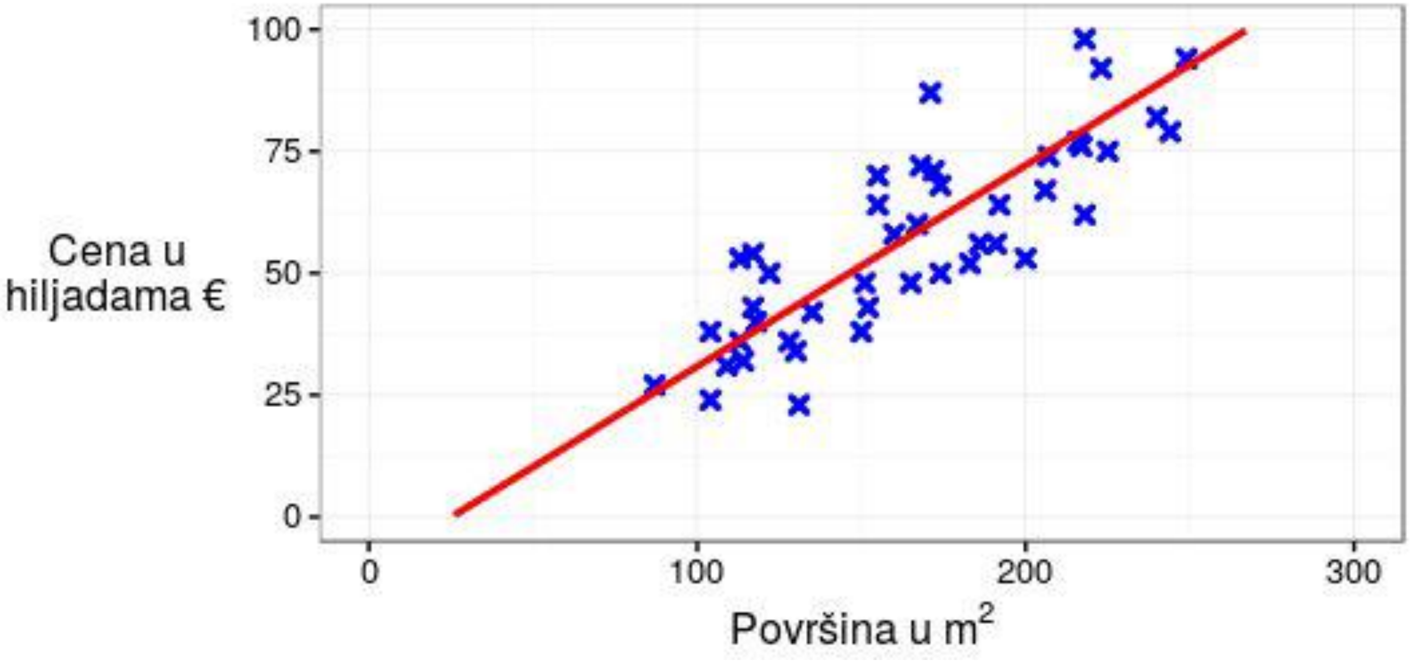

1. Strong, Positive
2. Weak, Positive
3. No correlation
4. Strong, Negative

Linearna regresija - primer

Predvideti **cenu** stana na osnovu njegove kvadrature.

Trening skup

Površina (u m^2)	Cena (u hiljadama €)
152	43
244	79
104	24
87	27
…	…



Linearna regresija

- Linearna regresija se koristi za predviđanje kontinualnih izlaznih vrednosti na osnovu ulaza.
- Linearna regresija modelira odnos dve veličine određivanjem linearne jednačine kojom su te dve veličine povezane.

Matematička notacija

Trening skup

Površina (u m^2)	Cena (u hiljadama €)
152	43
244	79
104	24
87	27
…	…

m - Broj trening primera

x - Ulazna promenljiva (kvadratura)

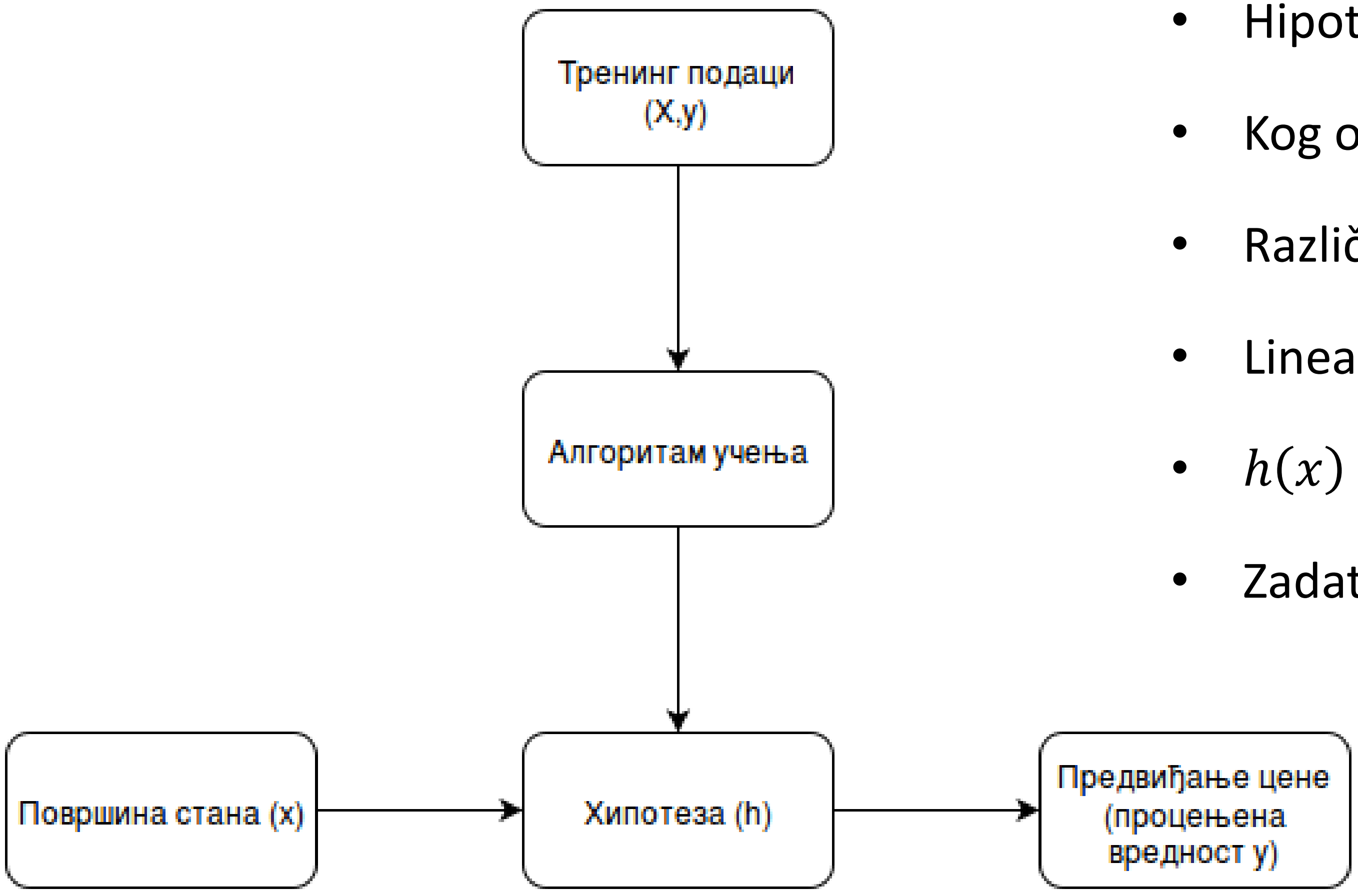
y - Izlazna promenljiva (cena)

(x, y) - Jedan trening primer, jedan red iz tabele

$(x^{(i)}, y^{(i)})$ - i-ti trening primer, i-ti red u tabeli

Npr. $x^{(1)} = 152,\ y^{(2)} = 79$

Hipoteza



- Hipoteza h - Funkcija koja mapira x u y
- Kog oblika je funkcija h?
- Različiti algoritmi – različiti oblici
- Linearna regresija - linearna funkcija
- $h(x) = w_0 + w_1 x$
- Zadatak algoritma – odrediti parametre w_0, w_1

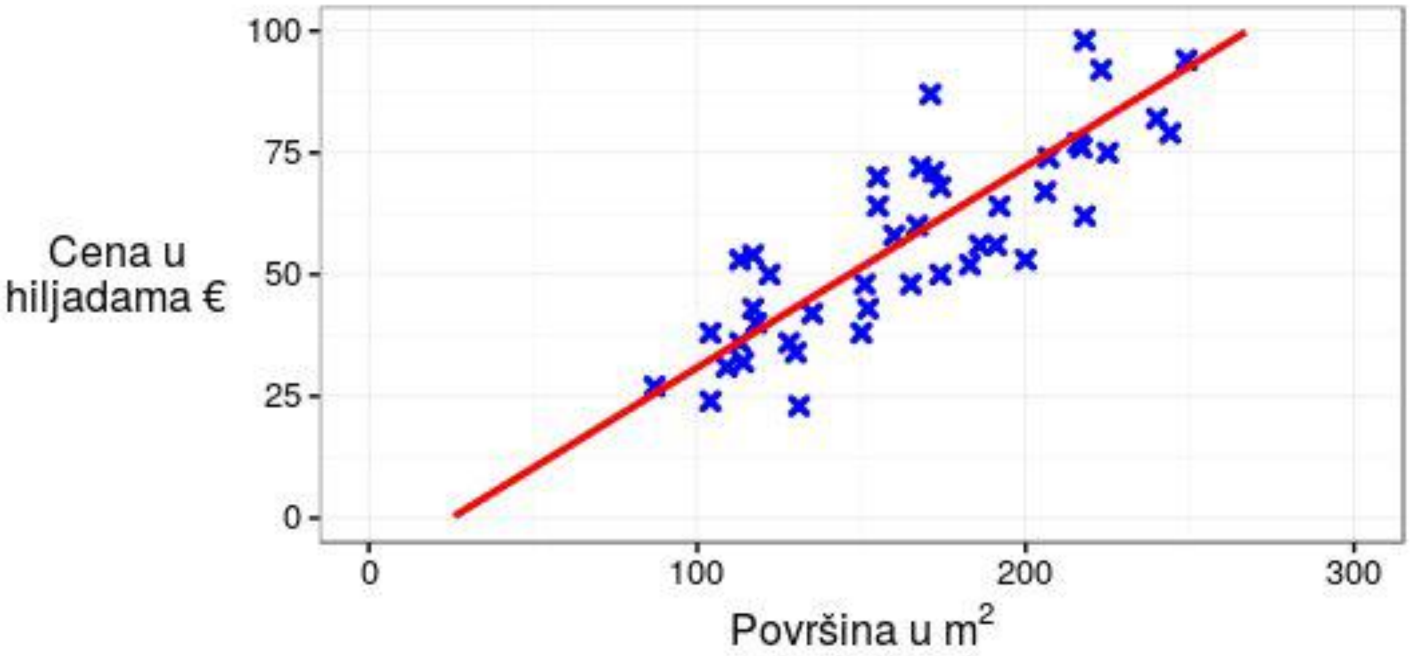
Hipoteza

Predvideti **cenu** stana na osnovu njegove kvadrature.

$$h(x) = w_0 + w_1 x$$

Trening skup

Površina (u m^2)	Cena (u hiljadama €)
152	43
244	79
104	24
87	27
…	…



Linearna regresija - osobine

- Linear models are parametric, meaning that they have a fixed form with a small number of numeric parameters that need to be learned from data. This is different from tree or rule models, where the structure of the model (e.g., which features to use in the tree, and where) is not fixed in advance.

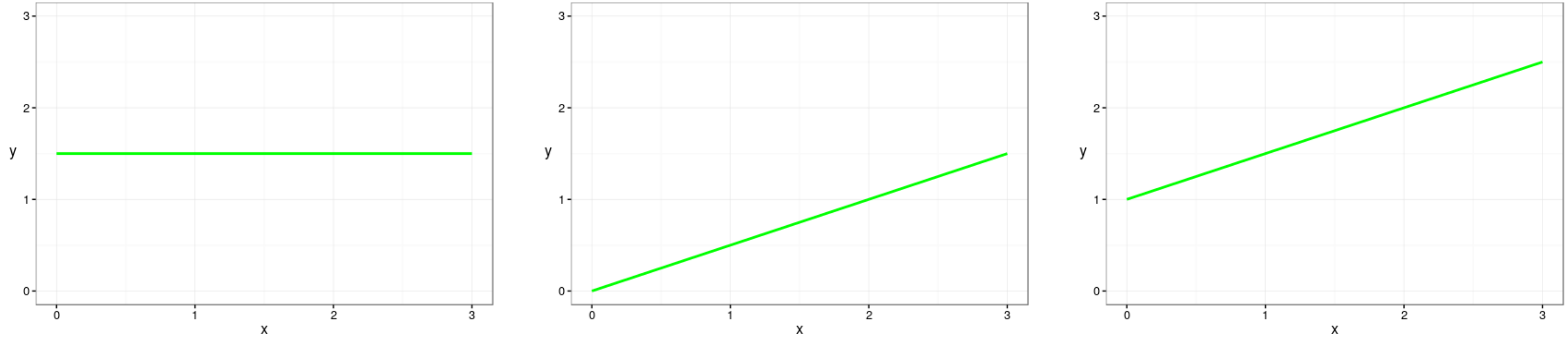
- Linear models are stable, which is to say that small variations in the training data have only limited impact on the learned model.

- Linear models have low variance but high bias – they tend to underfit the data.

Hipoteza

$$h(x) = w_0 + w_1 x$$

- Različiti parametri w_0 i w_1 daju različite hipoteze.



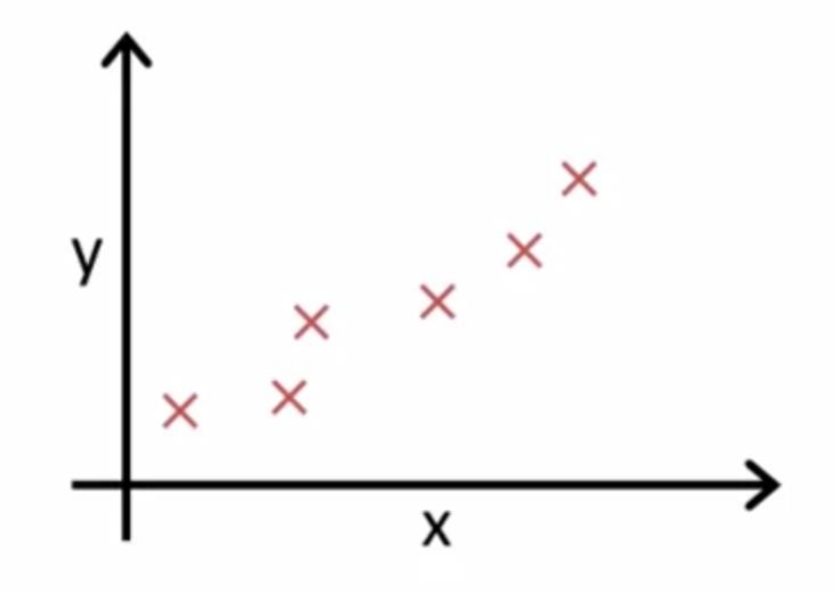
$w_0 = 1{,}5$
$w_1 = 0$

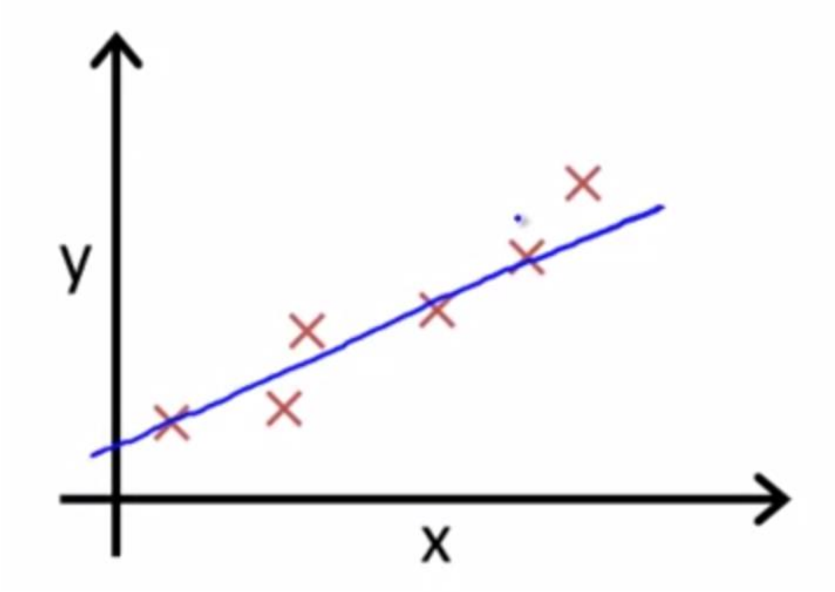
$w_0 = 0$
$w_1 = 0{,}5$

$w_0 = 1$
$w_1 = 0{,}5$

- Kako odabrati odgovarajuće vrednosti parametara modela w_0 i w_1?

Određivanje modela linearne regresije





- Određivanje modela linearne regresije svodi se na određivanje parametara w_0 i w_1, takvih da dobijena prava $h(x) = w_0 + w_1 x$ što bolje oslikava podatke iz trening skupa

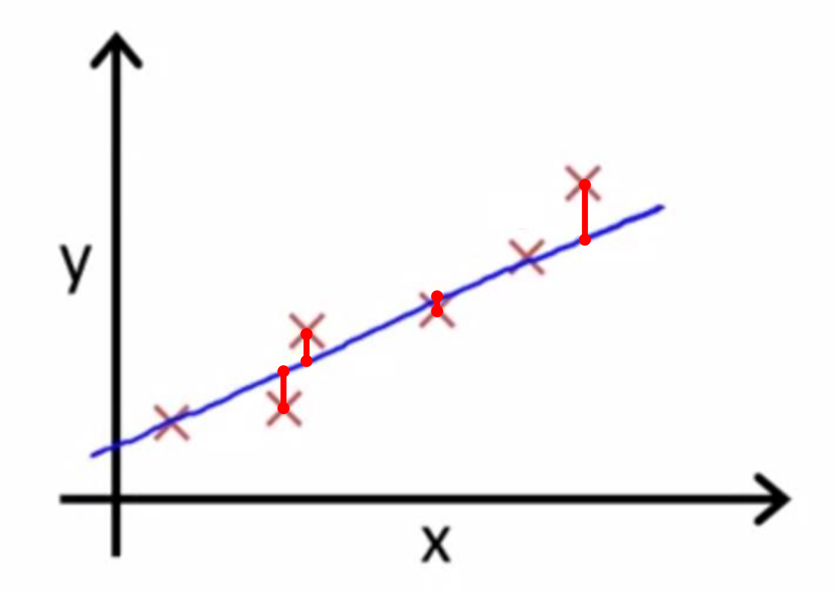
Određivanje modela linearne regresije

- Minimizacija razlika između $h(x)$ dobijenog modelom i konkretnih vrednosti y iz podataka.

$$\min_{w_0, w_1} (h(x) - y)$$

- Razlika $\hat{y}_i = h(x_i)$:

$$e_i = y_i - \hat{y}_i \text{ - rezidual}$$

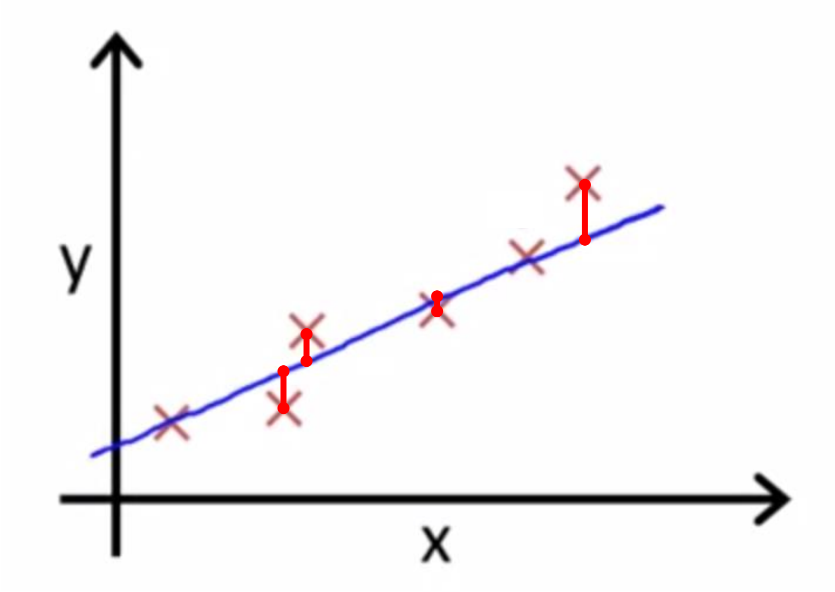


IMI 2019

Određivanje modela linearne regresije

- Minimizacija razlika između $y = h(x)$ dobijenog modelom i konkretnih vrednosti y iz podataka.

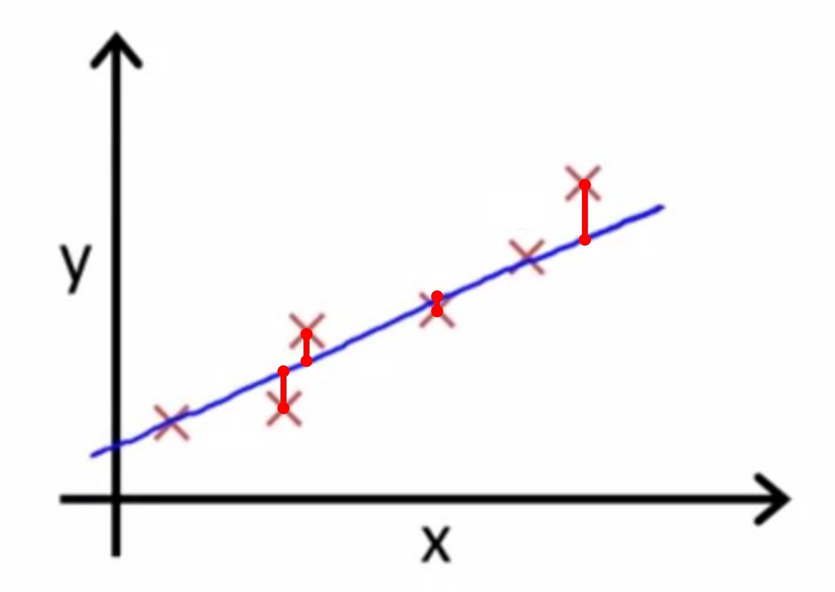
$$\min_{w_0, w_1} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right)$$



Određivanje modela linearne regresije

- Minimizacija razlika između $y = h(x)$ dobijenog modelom i konkretnih vrednosti y iz podataka.

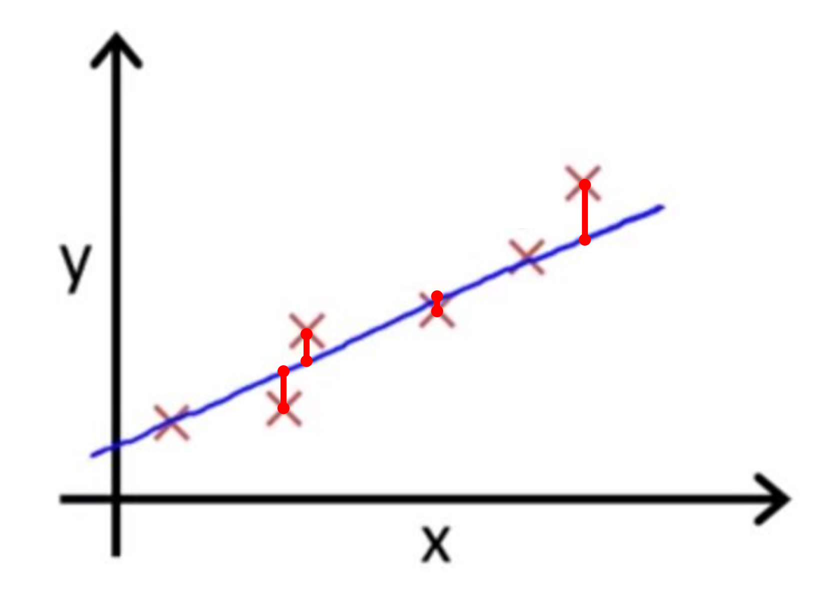
$$\min_{w_0, w_1} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$$



Određivanje modela linearne regresije

- Minimizacija razlika između $y = h(x)$ dobijenog modelom i konkretnih vrednosti y iz podataka.

$$\min_{w_0, w_1} \frac{1}{2m} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

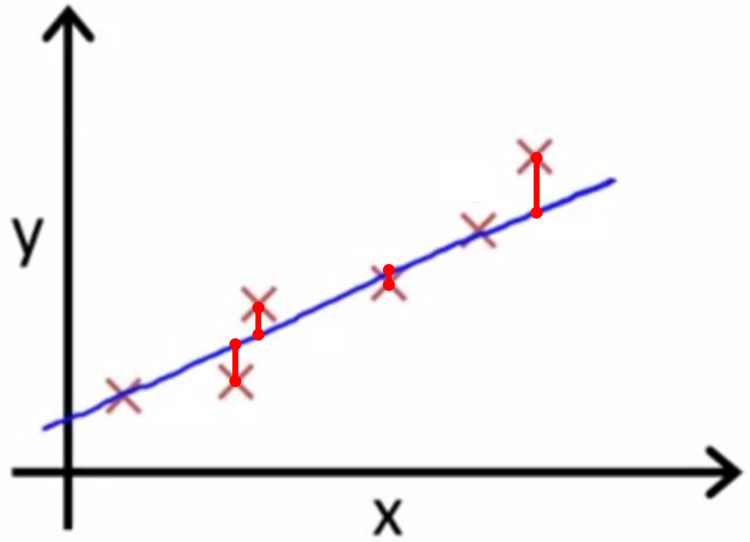


Funkcija ocene greške (*cost function*)

- Meri razliku između $h(x)$ i y, predviđanja i prave vrednosti.

- Funkcija **srednje kvadratne greške**

$$E(w_0, w_1) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$



Određivanje modela se svodi na odabir takvih parametara modela za koje model pravi najmanju grešku na trening skupu.

Funkcija ocene greške - vizuelizacija

- Hipoteza

$$h(x) = w_0 + w_1 x$$

- Parametri:

$$w_0, w_1$$

- Funkcija ocene greške:

$$E(w_0, w_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Cilj

$$\min_{w_0, w_1} E(w_0, w_1)$$

Funkcija ocene greške - vizuelizacija

- Hipoteza

$$h(x) = w_0 + w_1 x$$

- Parametri:

$$w_0, w_1$$

- Funkcija ocene greške:

$$E(w_0, w_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

- Cilj

$$\min_{w_0, w_1} E(w_0, w_1)$$

- Radi jednostavnijeg prikaza:

$$w_0 = 0 \;\rightarrow h(x) = w_1 x$$

- Parametar:

$$w_1$$

- Funkcija ocene greške:

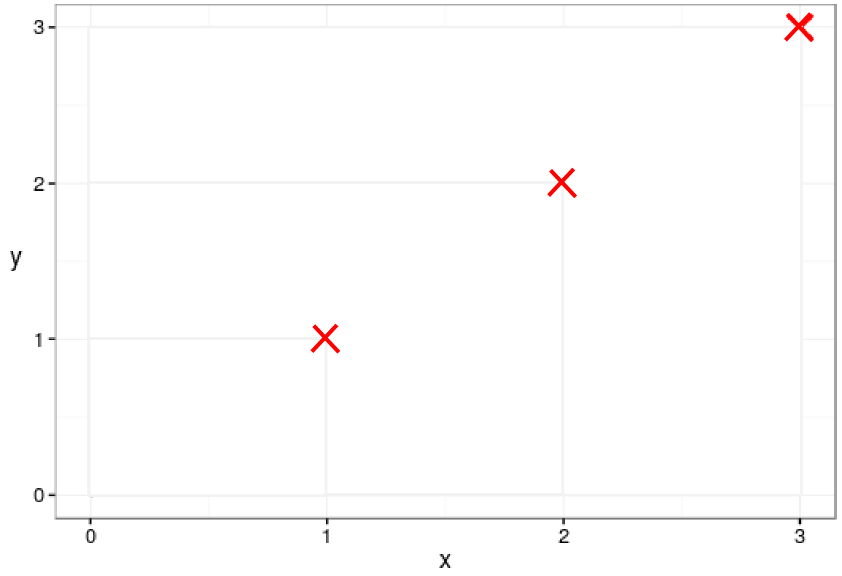
$$E(w_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(w_1 x^{(i)} - y^{(i)}\right)^2$$

- Cilj

$$\min_{w_1} E(w_1)$$

Funkcija ocene greške - vizuelizacija

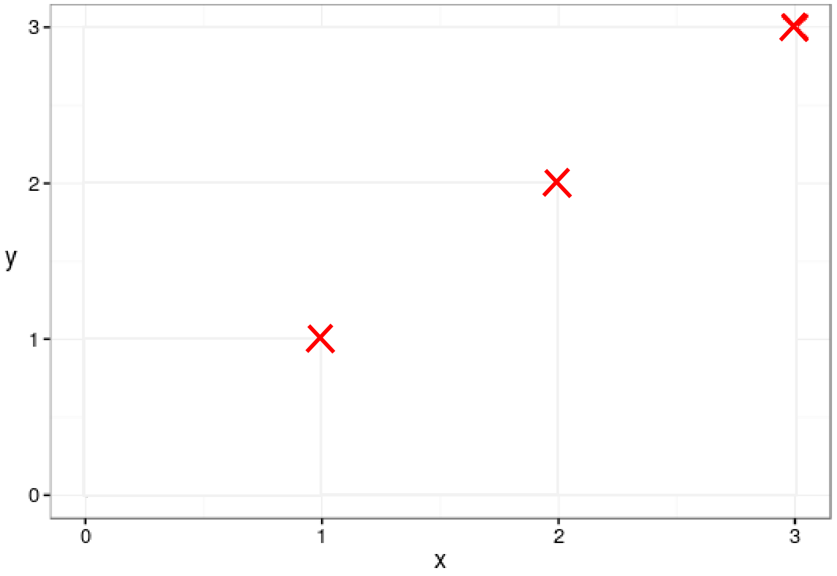
- $h(x)$ - funkcija od x
- $E(w_1)$ – funkcija od w_1



Trening skup

Funkcija ocene greške - vizuelizacija

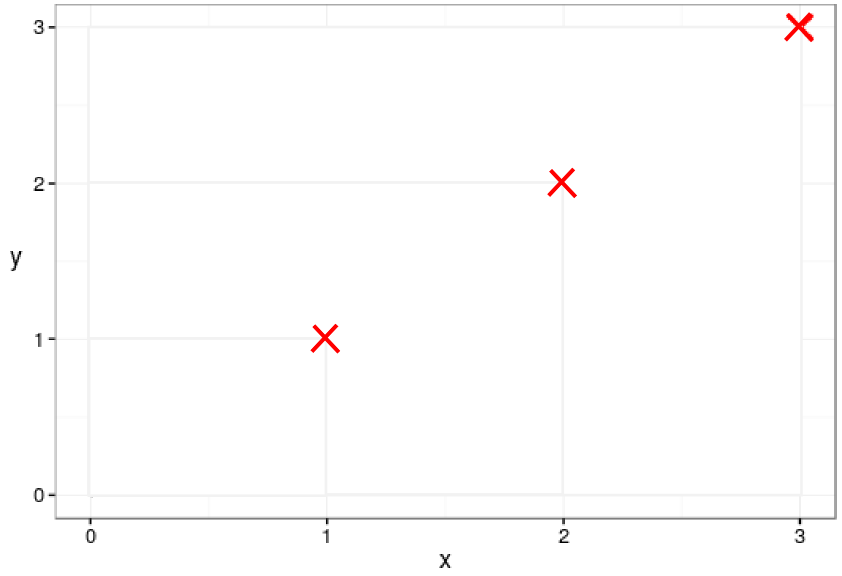
- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 1$



- $E(w_1)$ – funkcija od w_1

Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 1 \rightarrow h(x) = x$

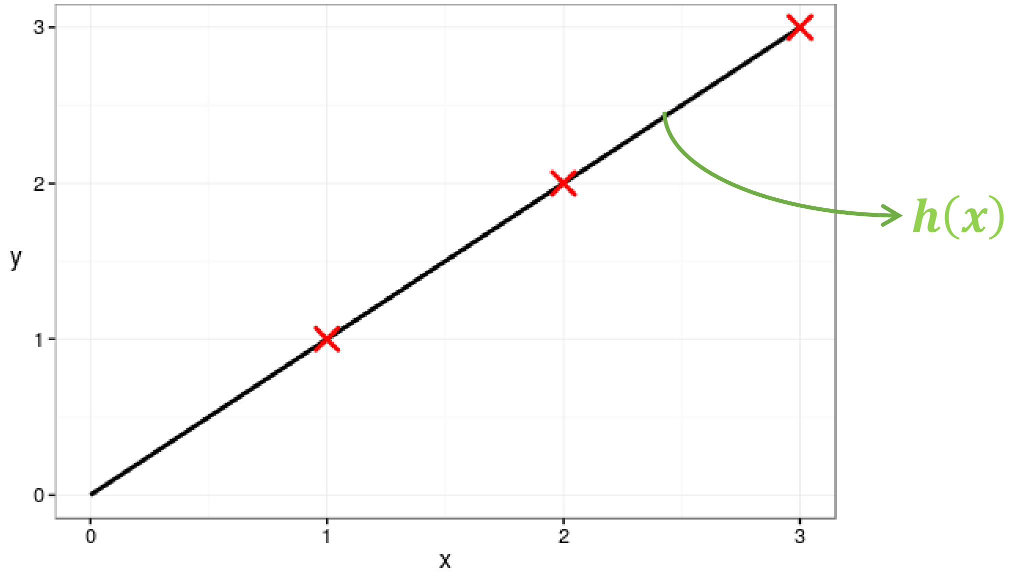


x	1	2	3
$h(x)$	1	2	3

- $E(w_1)$ – funkcija od w_1

Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 1 \rightarrow h(x) = x$

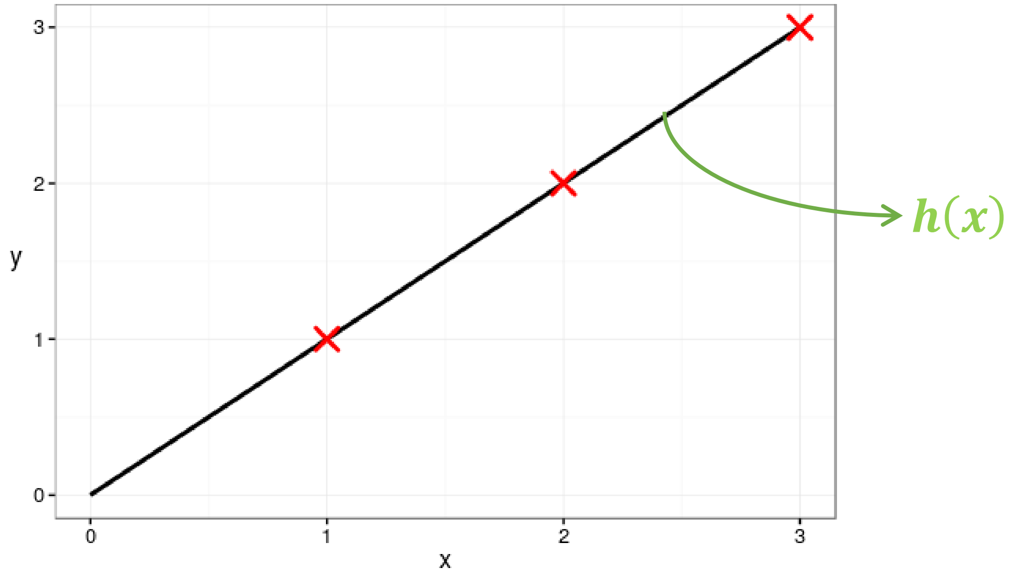


x	1	2	3
$h(x)$	1	2	3

- $E(w_1)$ – funkcija od w_1

Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 1 \rightarrow h(x) = x$

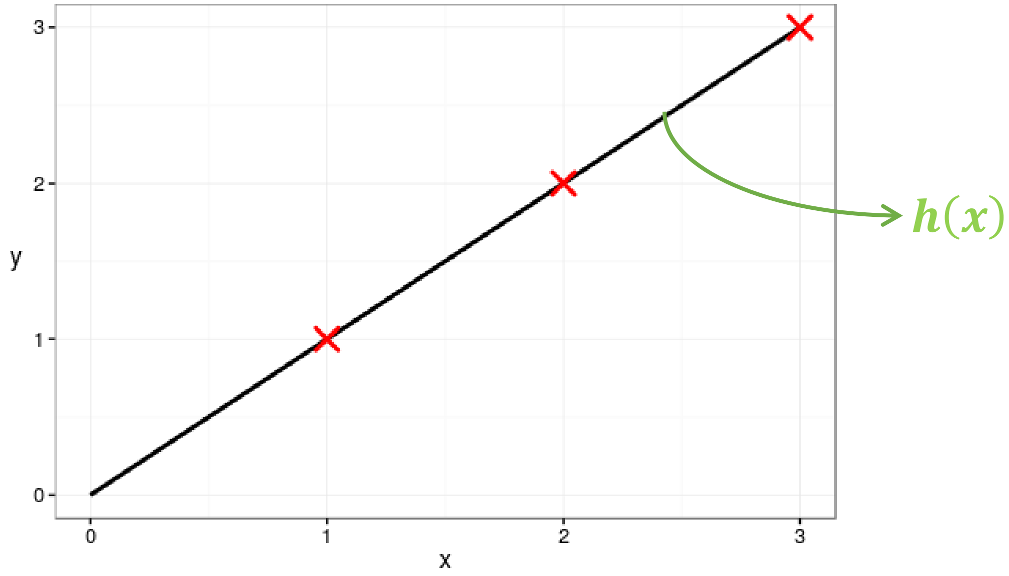


x	1	2	3
$h(x)$	1	2	3

- $E(w_1)$ – funkcija od w_1
- Kolika je funkcija ocene greške u slučaju kada je $w_1 = 1$?

Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 1 \rightarrow h(x) = x$



x	1	2	3
$h(x)$	1	2	3

- $E(w_1)$ – funkcija od w_1
- Kolika je funkcija ocene greške u slučaju kada je $w_1 = 1$?

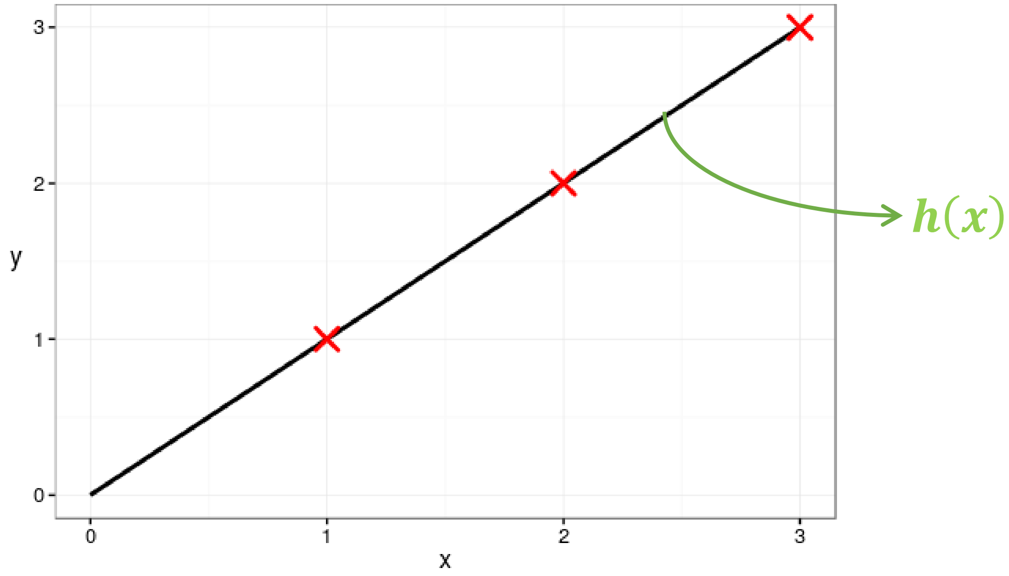
$$E(w_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{1}{2m}\sum_{i=1}^{m}\left(w_1 x^{(i)} - y^{(i)}\right)^2 =$$

$$= \frac{1}{2m}(0^2 + 0^2 + 0^2) = 0$$

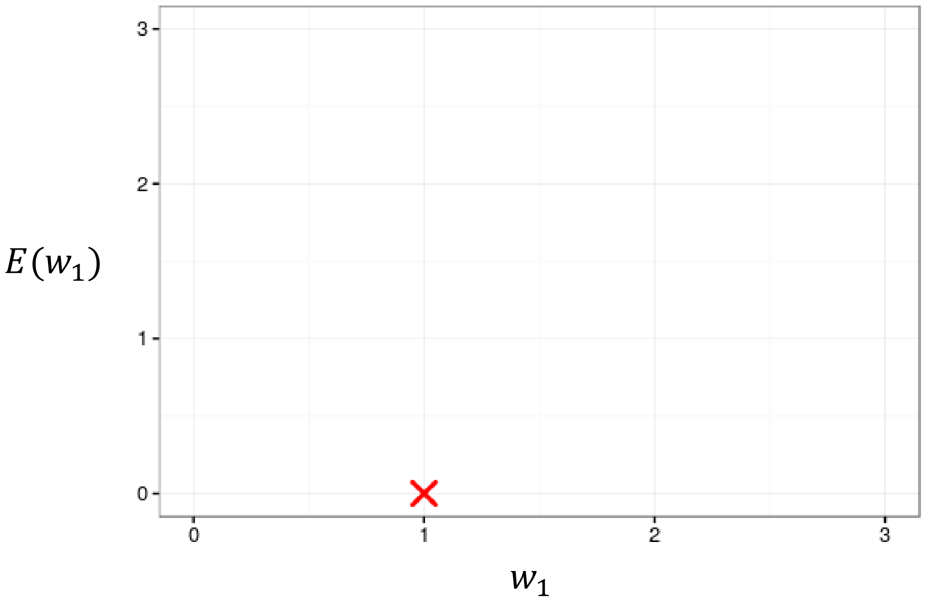
Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 1 \rightarrow h(x) = x$



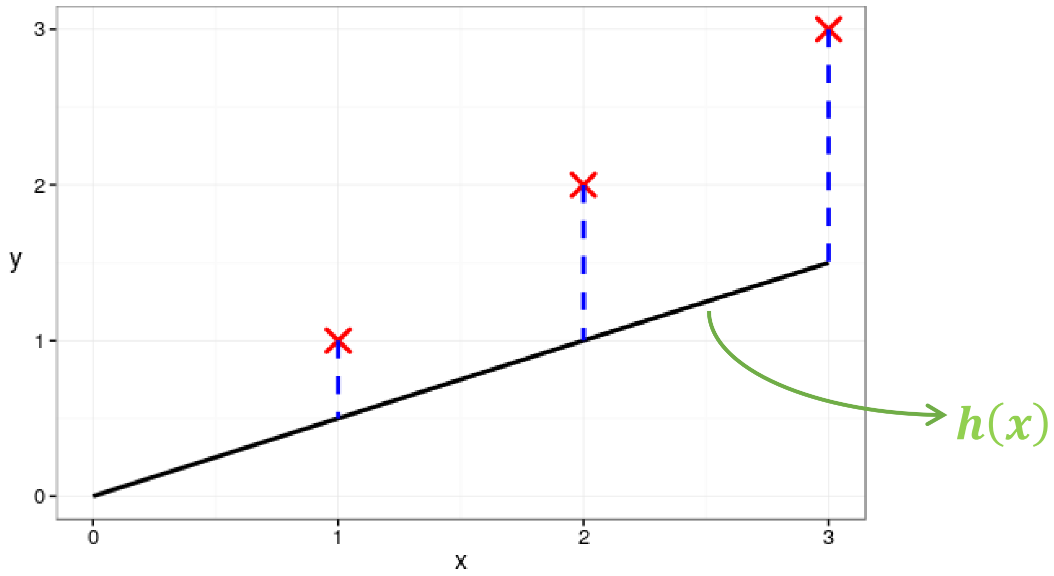
x	1	2	3
$h(x)$	1	2	3

- $E(w_1)$ – funkcija od w_1
- $E(w_1) = 0$



Funkcija ocene greške - vizuelizacija

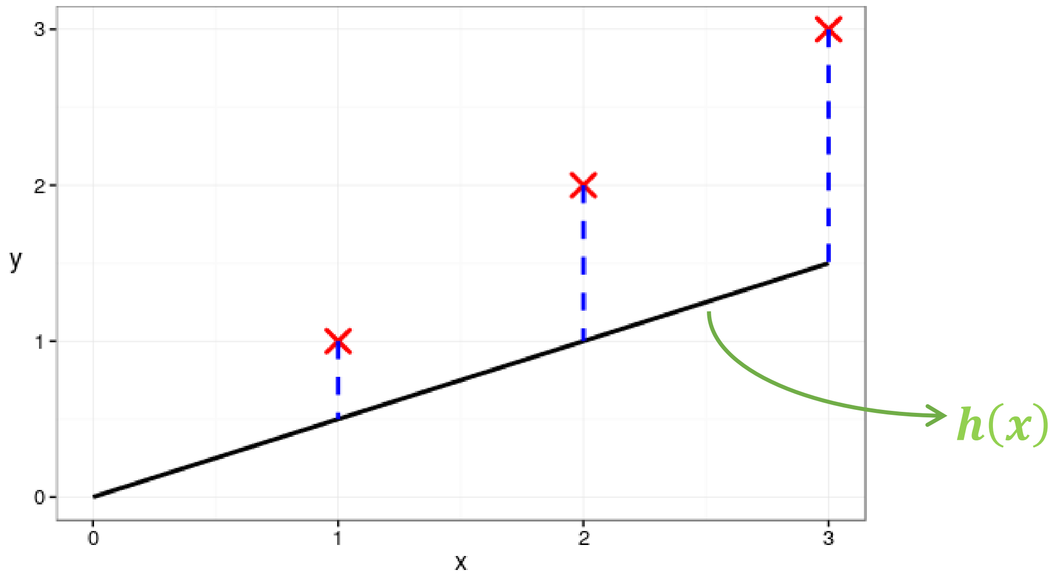
- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 0{,}5 \rightarrow h(x) = 0{,}5x$



x	1	2	3
$h(x)$	0,5	1	1,5

Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 0{,}5 \rightarrow h(x) = 0{,}5x$

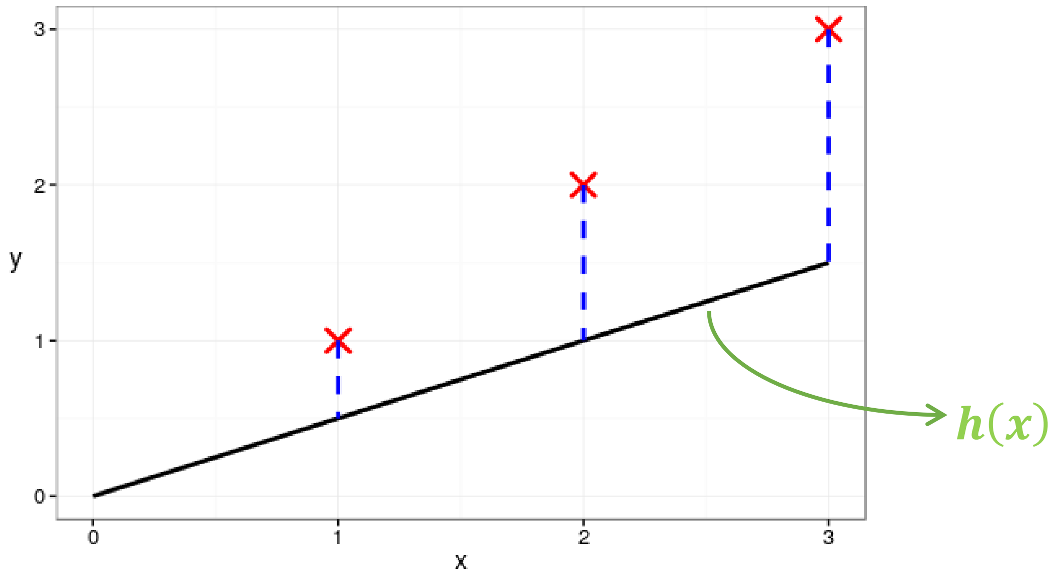


x	1	2	3
$h(x)$	0,5	1	1,5

- $E(w_1)$ – funkcija od w_1
- $E(0{,}5) = \frac{1}{2m}[(0{,}5-1)^2 + (1-2)^2 + (1{,}5-3)^2] = 0{,}58$

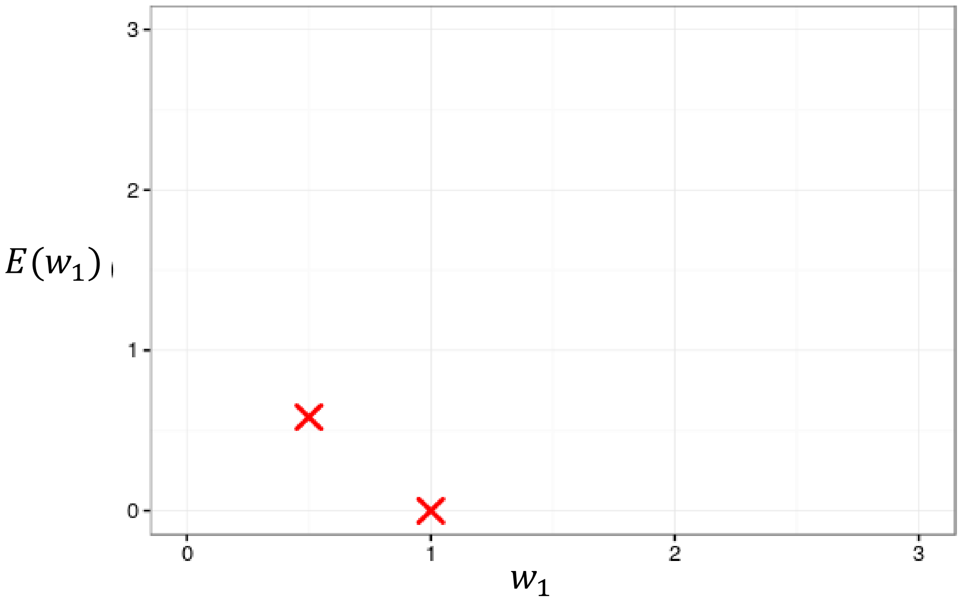
Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 0{,}5 \rightarrow h(x) = 0{,}5x$



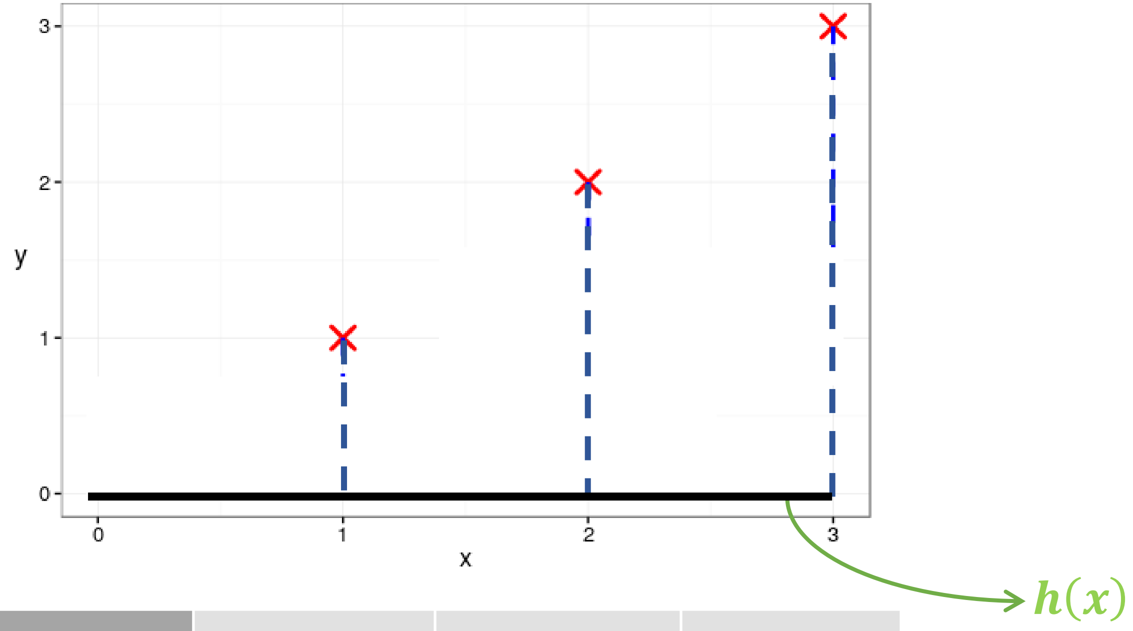
x	1	2	3
$h(x)$	0,5	1	1,5

- $E(w_1)$ – funkcija od w_1
- $E(0{,}5) = \frac{1}{2m}[(0{,}5-1)^2 + (1-2)^2 + (1{,}5-3)^2] = 0{,}58$



Funkcija ocene greške - vizuelizacija

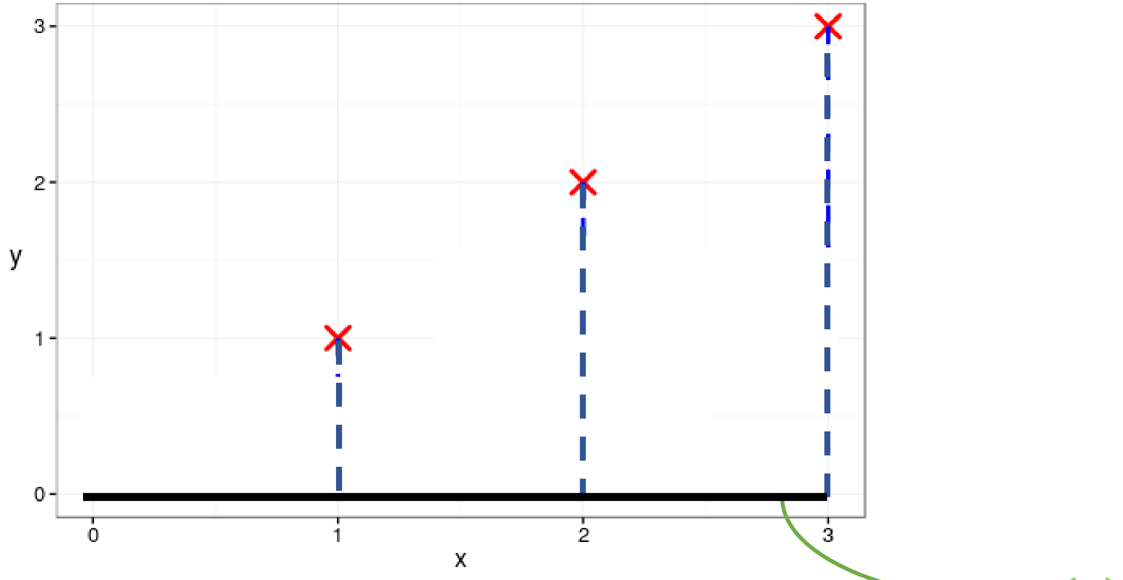
- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 0 \rightarrow h(x) = 0$



x	1	2	3
$h(x)$	0	0	0

Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 0 \rightarrow h(x) = 0$

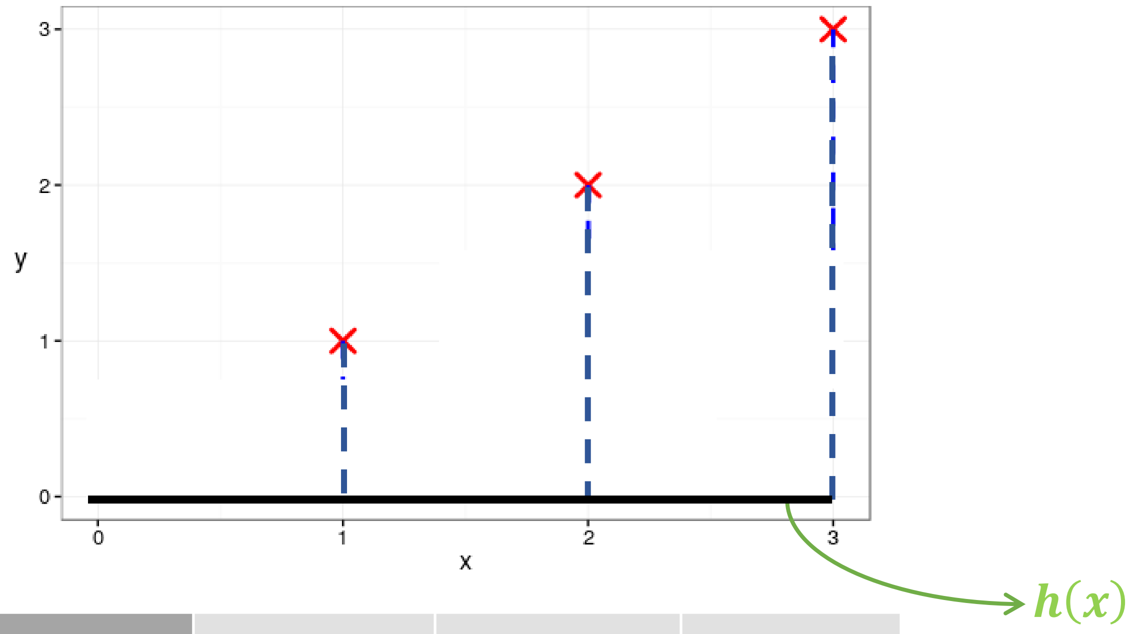


x	1	2	3
$h(x)$	0	0	0

- $E(w_1)$ – funkcija od w_1
- $E(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2] = 2{,}3$

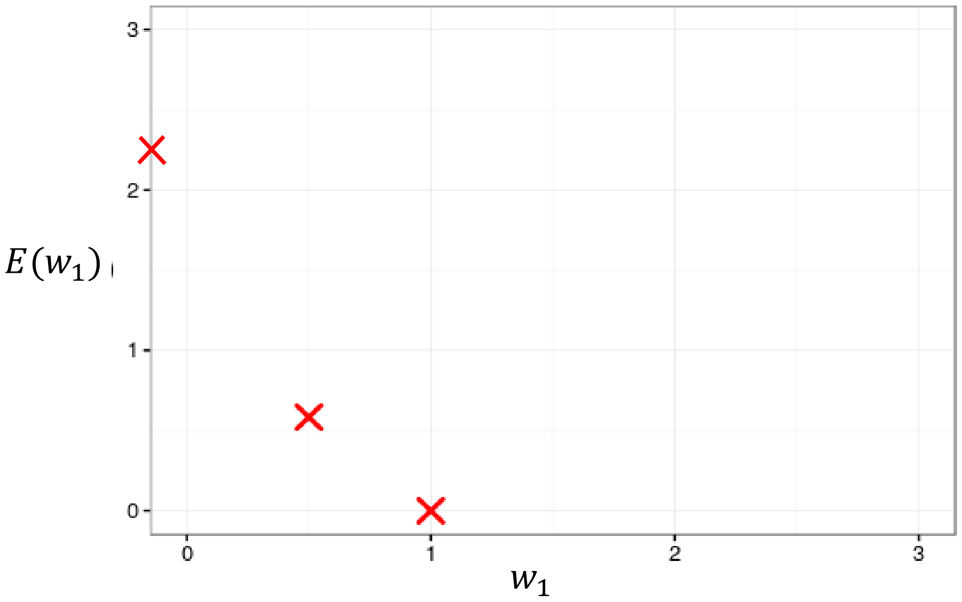
Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Uzmimo da je $w_1 = 0 \rightarrow h(x) = 0$



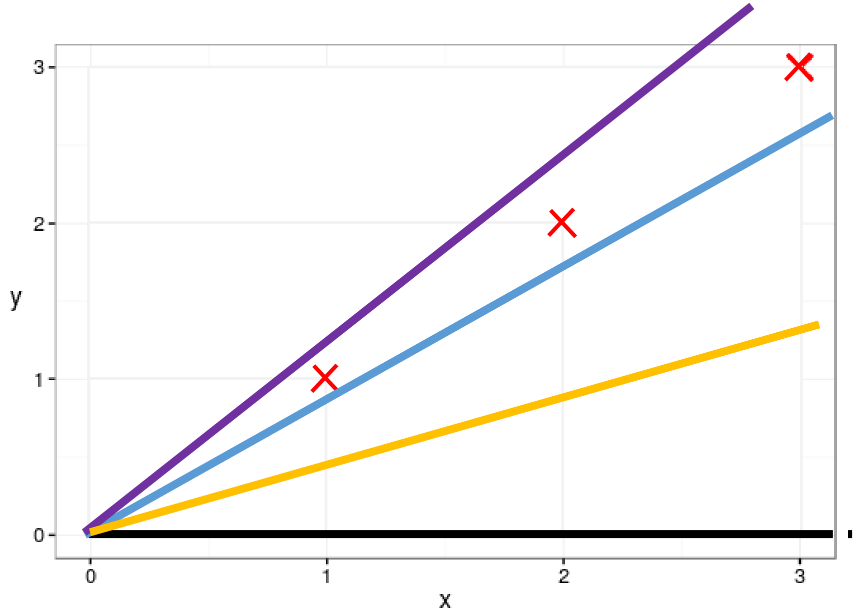
x	1	2	3
$h(x)$	0	0	0

- $E(w_1)$ – funkcija od w_1
- $E(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2] = 2{,}3$

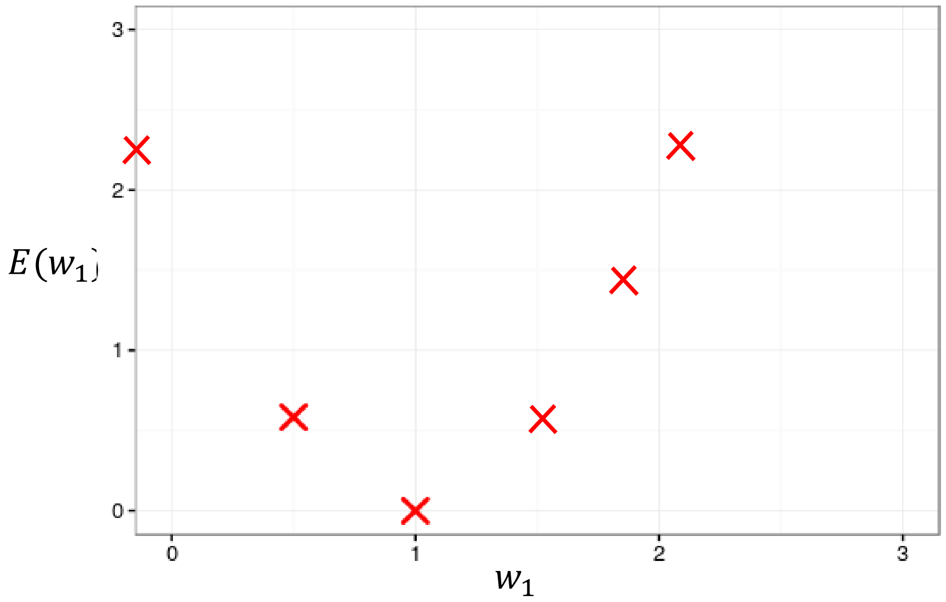


Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Za različite vrednosti w_1 dobićemo različite hipoteze

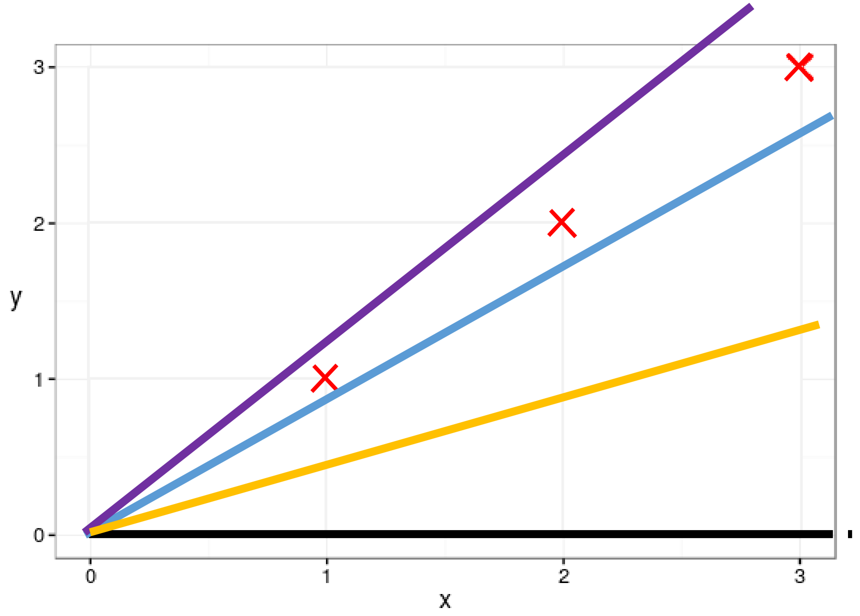


- $E(w_1)$ – funkcija od w_1
- Za različite vrednosti w_1 dobićemo različite vrednosti funkcije ocene greške

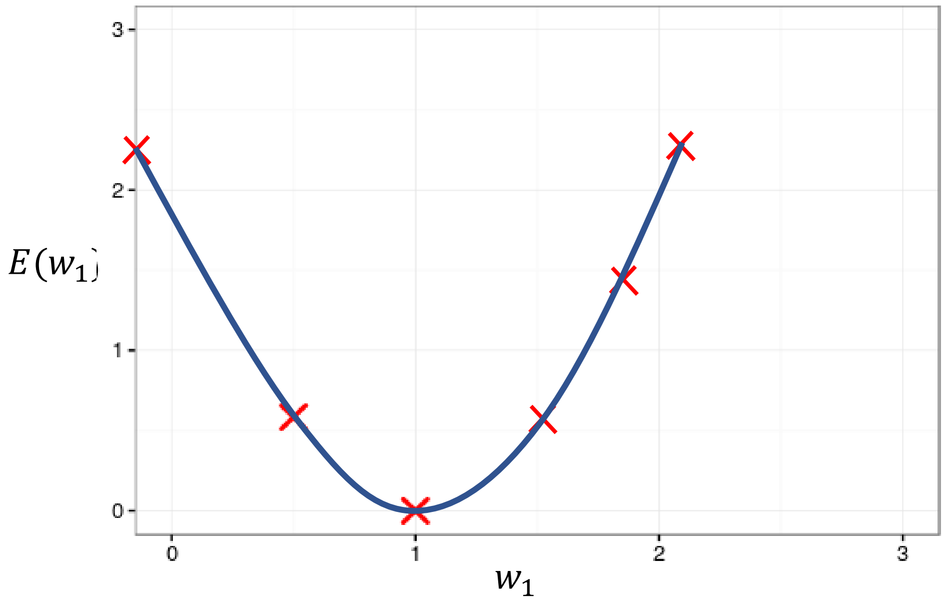


Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Za različite vrednosti w_1 dobićemo različite hipoteze

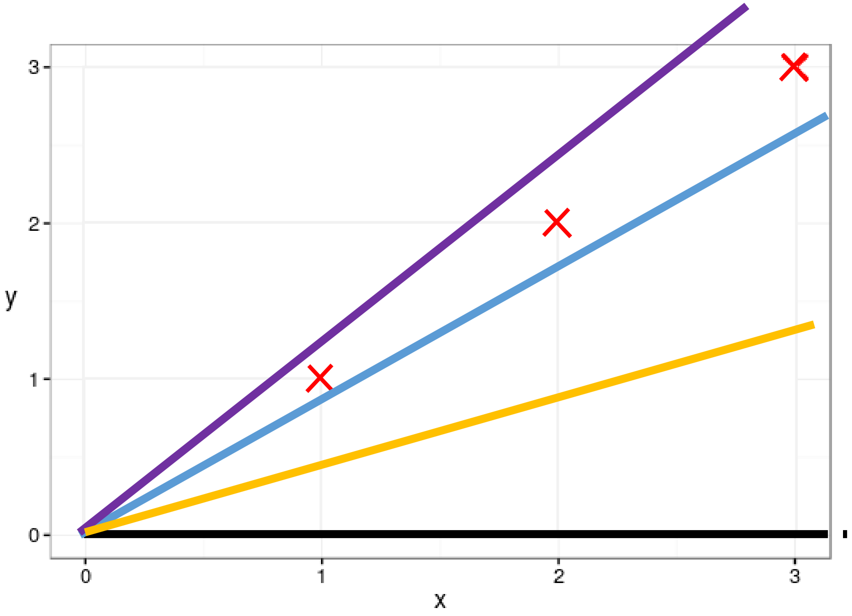


- $E(w_1)$ – funkcija od w_1
- Za različite vrednosti w_1 dobićemo različite vrednosti funkcije ocene greške

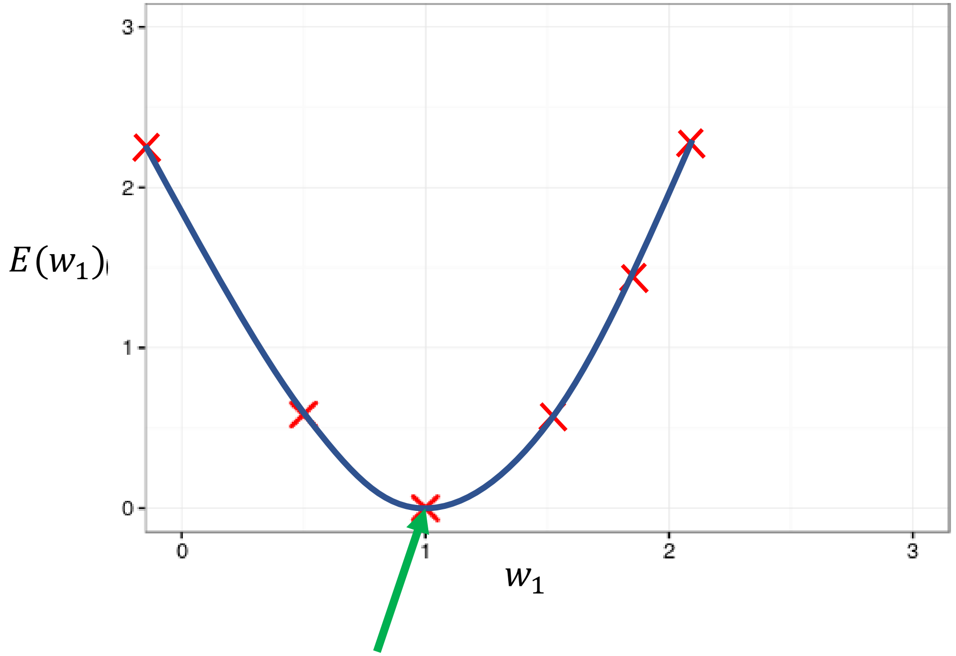


Funkcija ocene greške - vizuelizacija

- $h(x)$ - funkcija od x
- Za različite vrednosti w_1 dobićemo različite hipoteze



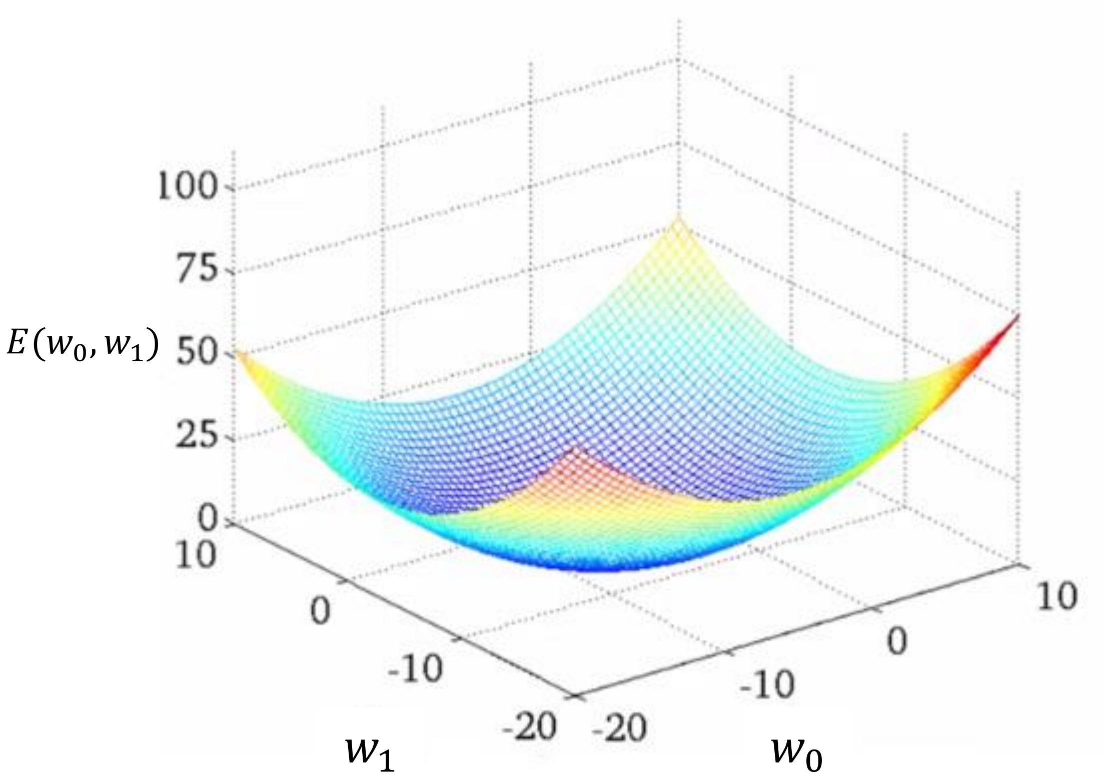
- $E(w_1)$ – funkcija od w_1
- Za različite vrednosti w_1 dobićemo različite vrednosti funkcije ocene greške



Biramo w_1 za koje funkcija ocene greške postiže minimum

Funkcija ocene greške - vizuelizacija

- Ako uključimo oba parametra - w_0, w_1
- Funkcija - $E(w_0, w_1)$ je sada površ
- Tražimo minimum ove funkcije



Minimum f-je greške

- Odrediti izvod funkcije greške po obe promenljive

$$\frac{\partial E}{\partial w_0} = \frac{1}{m}\sum_{i=1}^{m}(w_1 x_i + w_0 - y_i)$$

$$\frac{\partial E}{\partial w_1} = \frac{1}{m}\sum_{i=1}^{m}((w_1 x_i + w_0 - y_i)\, x_i)$$

Minimum f-je greške

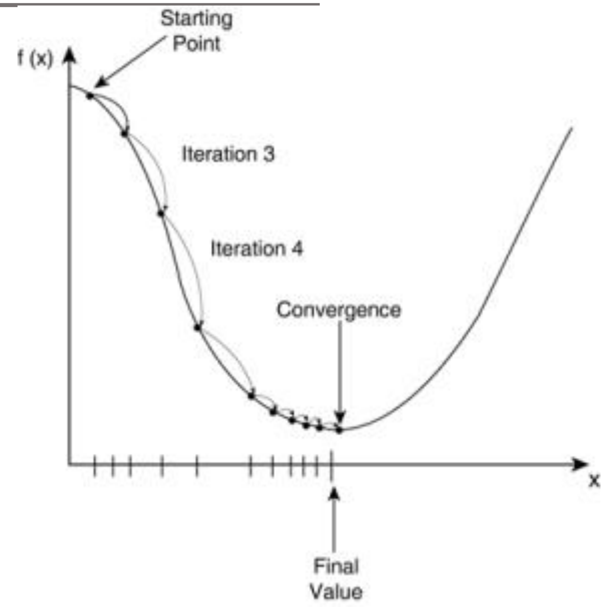
- Funkcija dostiže minimum za one vrednosti w_0, w_1 za koje su vrednosti parcijalnih izvoda jednake 0:

$$w_1 = \frac{m\left(\sum_{i=1}^{m} x_i y_i\right) - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} y_i\right)}{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}$$

$$w_0 = \frac{\left(\sum_{i=1}^{m} y_i\right) - w_1\left(\sum_{i=1}^{m} x_i\right)}{m}$$

Opadajući gradijent

- Imamo funkciju $E(w_0, w_1)$
- Želimo $\min\limits_{w_0, w_1} E(w_0, w_1)$
- Ideja
  - Krenemo od početnih w_0, w_1 (recimo $w_0 = 1, w_1 = 1$)
  - Korak po korak menjamo w_0, w_1 tako da smanjujemo $E(w_0, w_1)$
  - Dok ne stignemo do minimuma funkcije



- Kako odrediti smer i dužinu koraka?

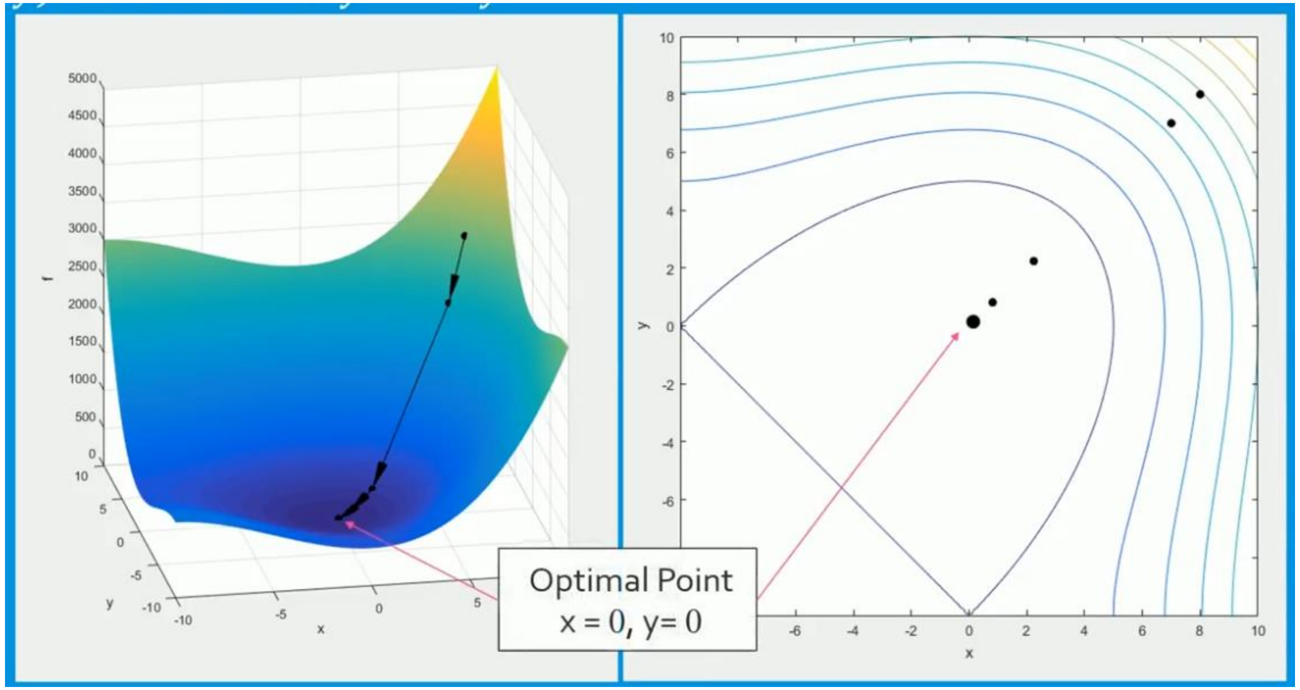
Gradijent

- Gradijent – izvod funkcije više promenljivih, prostorni izvod.

$$\nabla E(w_0, w_1) = \left(\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}\right)$$

- Parcijalni izvod funkcije po nekoj promenljivoj pokazuje koliko se brzo menja funkcija po toj promenljivoj

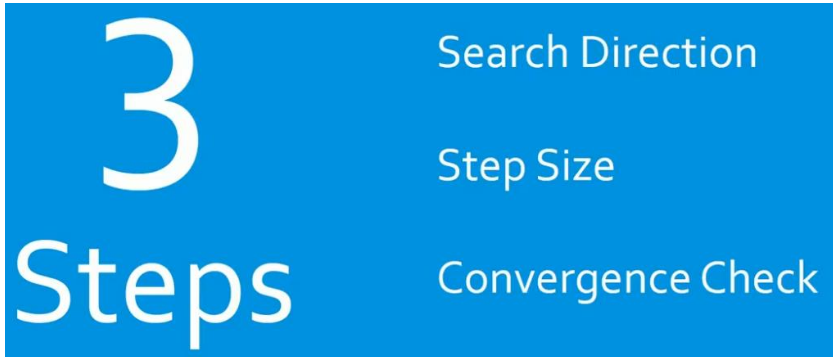
- Smer vektora gradijenta je smer u kom funkcija najviše raste.



Gradijentni spust

U svakoj iteraciji se pomeramo u smeru suprotnom od smera gradijenta.

$$w_j^{(k+1)} = w_j^{(k)} - \alpha_k \frac{\partial}{\partial w_j} E\left(w_0^{(k)}, w_1^{(k)}\right), \mathrm{j} = 1{,}2$$



α_k - veličina koraka (**stopa učenja**)

Različiti načini izbora koraka za spust, daju različite algoritme zasnovane na gradijentu.

Opadajući gradijent

- Ispravno: istovremena promena

$$temp_0 = w_0^{(k)} - \alpha_k \frac{\partial}{\partial w_0} E\left(w_0^{(k)}, w_1^{(k)}\right)$$

$$temp_1 = w_1^{(k)} - \alpha_k \frac{\partial}{\partial w_1} E\left(w_0^{(k)}, w_1^{(k)}\right)$$

$$w_0 = temp_0$$

$$w_1 = temp_1$$

- Pogrešno

$$temp_0 = w_0^{(k)} - \alpha_k \frac{\partial}{\partial w_0} E\left(w_0^{(k)}, w_1^{(k)}\right)$$

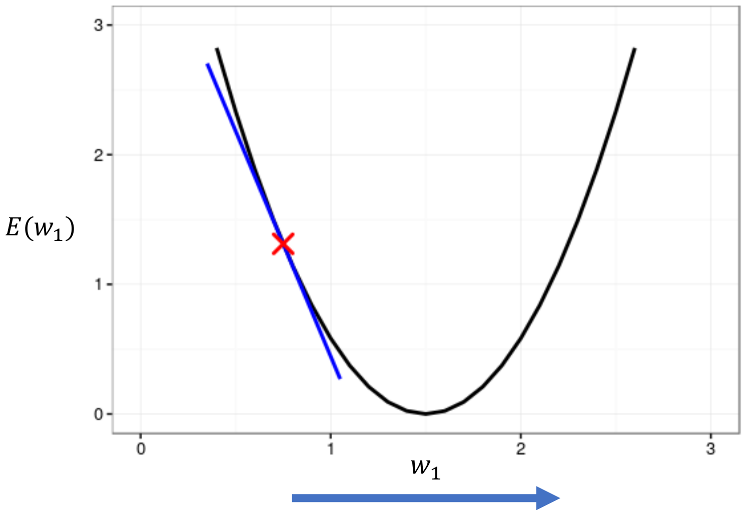
$$w_0 = temp_0$$

$$temp_1 = w_1^{(k)} - \alpha_k \frac{\partial}{\partial w_1} E\left(w_0^{(k)}, w_1^{(k)}\right)$$

$$w_1 = temp_1$$

Opadajući gradijent

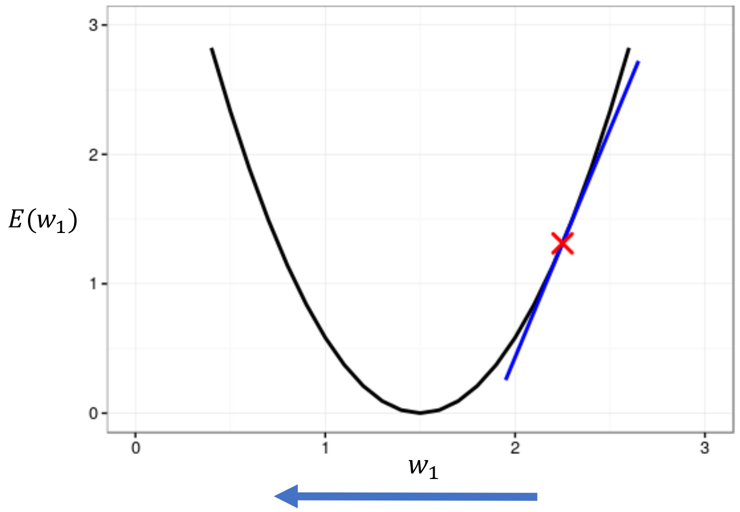
- Izvod 1D funkcije
- $w_1 = w_1 - \alpha \cdot E'(w_1)$



$$E'(w_1) < 0$$

$$w_1 = w_1 - \alpha \cdot E'(w_1)$$

w_1 raste



$$E'(w_1) > 0$$

$$w_1 = w_1 - \alpha \cdot E'(w_1)$$

w_1 opada

Opadajući gradijent za linearnu regresiju

Konačan oblik algoritma:

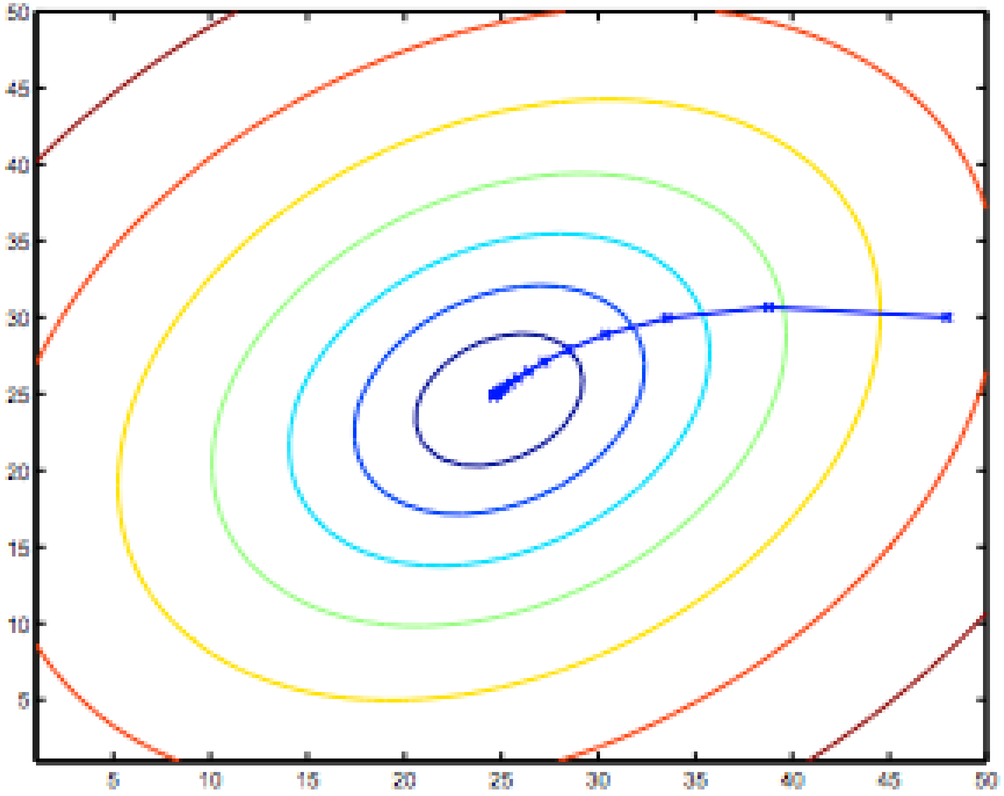
Repeat until convergence {

$$w_0 = w_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)$$

$$w_1 = w_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} ((h_\theta(x_i) - y_i)\, x_i)$$

}

w_0 i w_1 menjaju vrednost istovremeno.



Uslov za zaustavljanje algoritma

Algoritam staje kada je $\nabla E\left(w_o^{(k+1)}, w_1^{(k+1)}\right) = 0.$

Ovaj uslov je u praksi kod iterativnih metoda retko zadovoljen, pa se koristi jedan od sledećih uslova:

$$\left|E\left(w_o^{(k+1)}, w_1^{(k+1)}\right) - E\left(w_o^{(k)}, w_1^{(k)}\right)\right| \leq \varepsilon$$

$$\left|w_i^{(k+1)} - w_i^{(k)}\right| \leq \varepsilon, i = 1,2$$