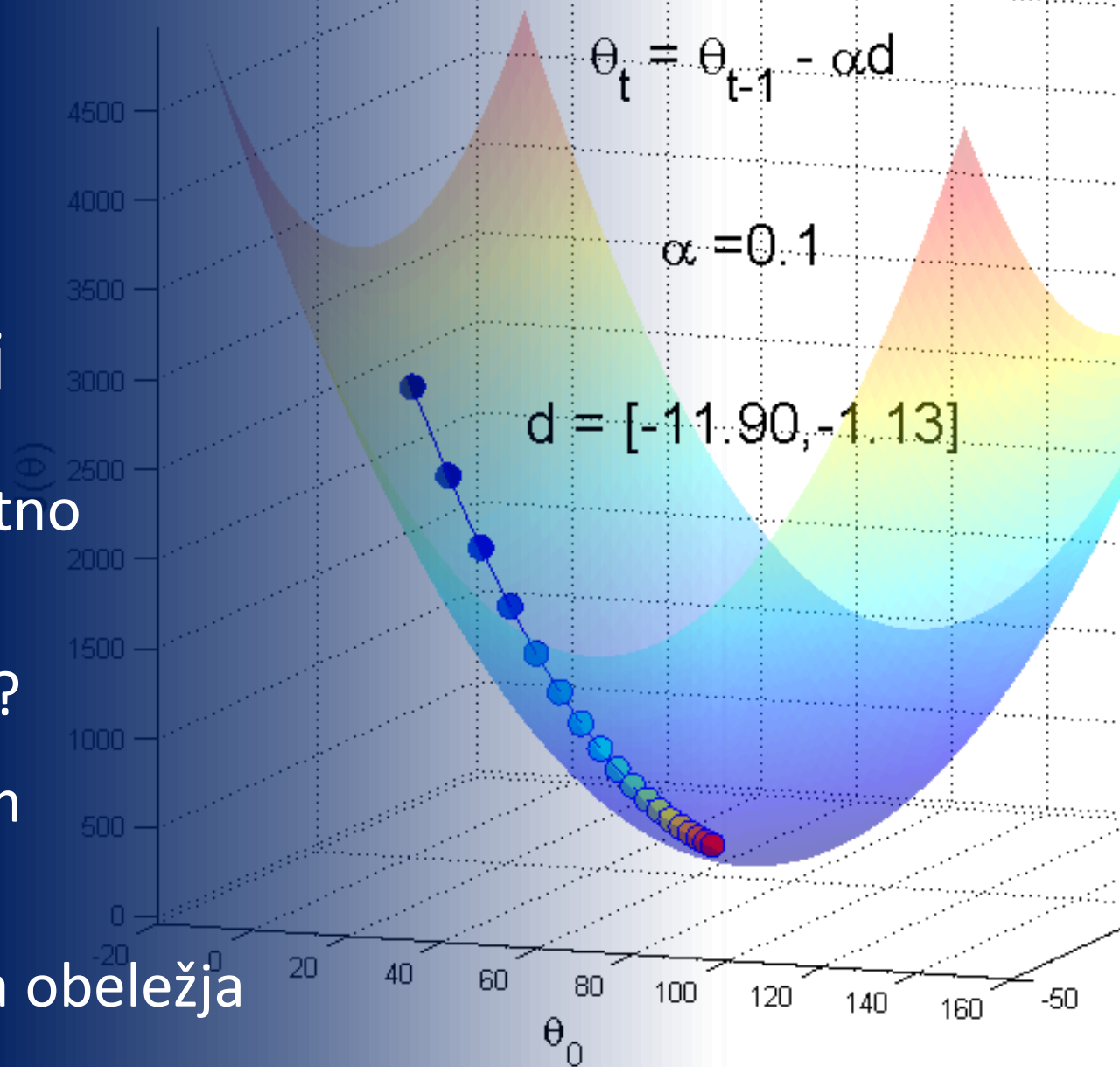


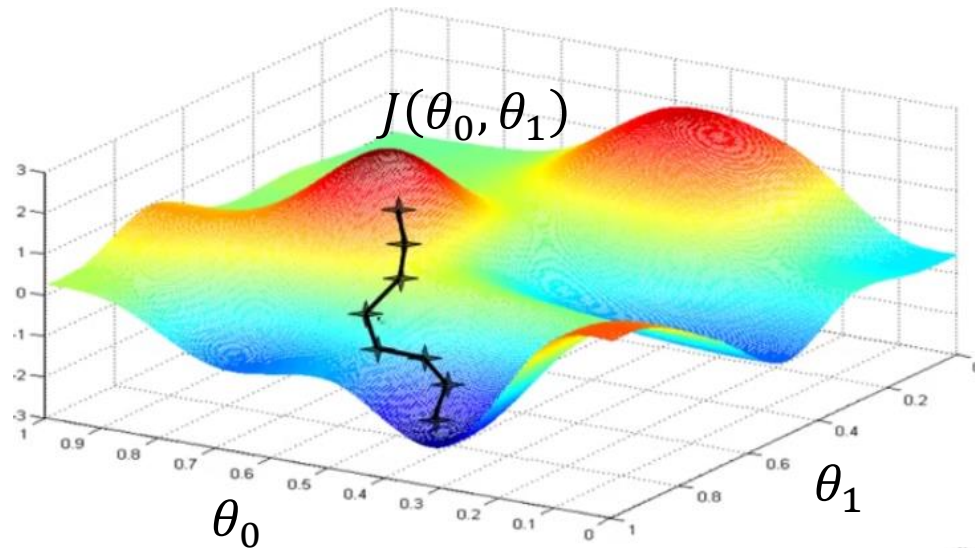
Gradijentni spust u praksi

- Kako utiče početno rešenje?
- Kako odabrati α ?
- Da li je algoritam konvergirao?
- Značaj skaliranja obeležja

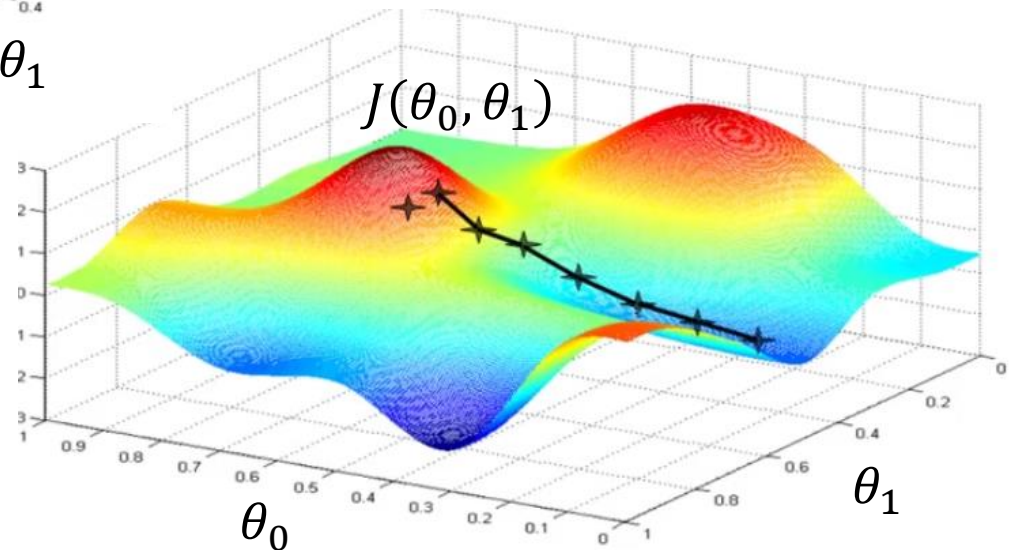


Odabir početnog rešenja $\theta^{(0)}$

Gradient descent pronalazi **lokalni** optimum



U zavisnosti od
inicijalizacije možemo
završiti u različitim tačkama

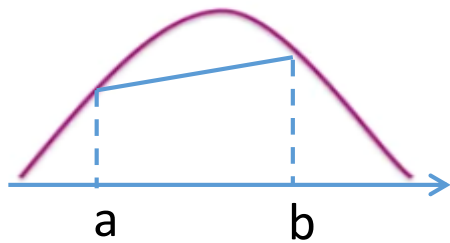


Da li je ovo
problem?

Da li *Gradient Descent* u slučaju
linearne regresije sa RSS merom
greške može da zaglavi u
lokalnom minimumu?

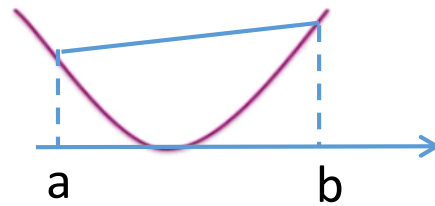
Odabir početnog rešenja θ u linearnoj regresiji

Konkavna funkcija



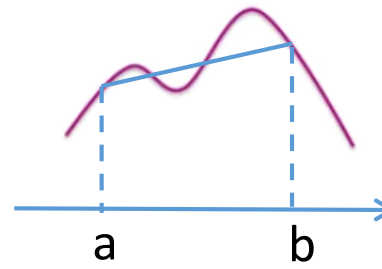
Jedan maksimum

Konveksna funkcija

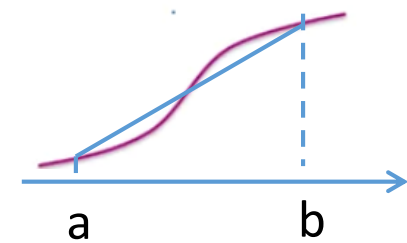


Jedan minimum

Ostale funkcije



Više ekstrema



Nema ekstrema

Za RSS grešku u linearnoj regresiji $J(\theta)$ je konveksno

Odabir α

LMS (*Least Mean Squares*) update rule

$$\theta_d^{(t+1)} = \theta_d^{(t)} - \overset{\text{Konstanta}}{\frac{\alpha}{N}} \sum_{i=1}^N \left(h_{\theta}(x^{(i)}) - y^{(i)} \right) x_d^{(i)}$$

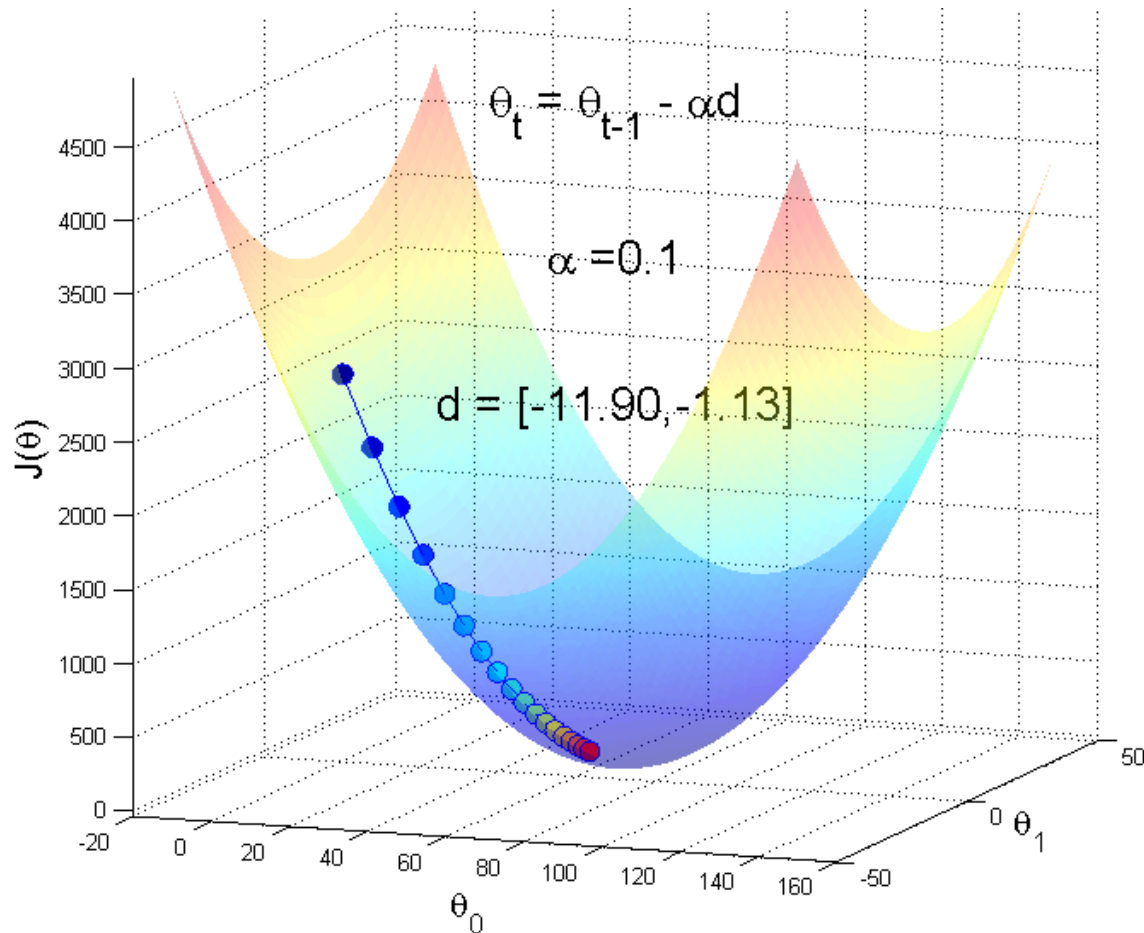
Promena θ je
proporcionalna grešci

Mala greška \rightarrow
nećemo mnogo menjati θ

Velika greška \rightarrow
mnogo ćemo promeniti θ

Što smo bliže minimumu, pravimo manje korake

Odabir α : konvergencija



Gradient descent može da konvergira u minimum za fiksirno α

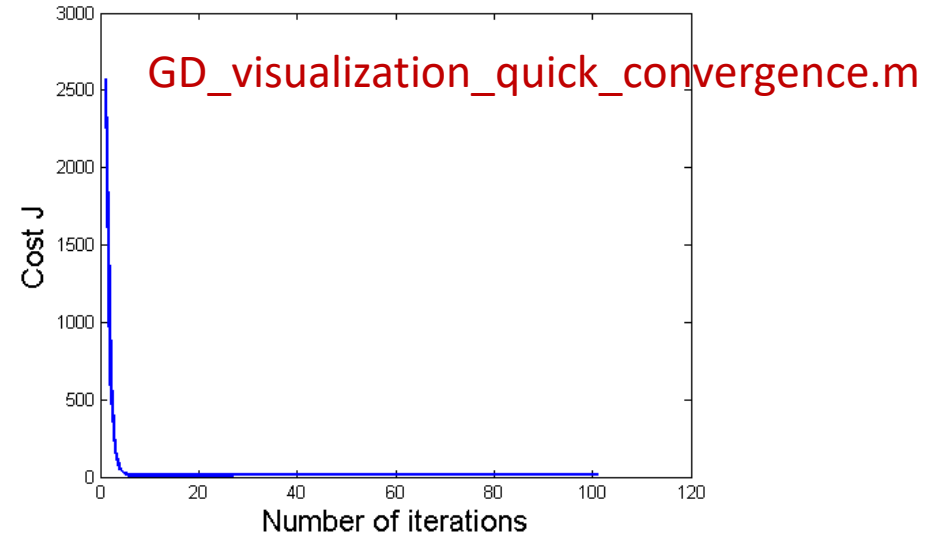
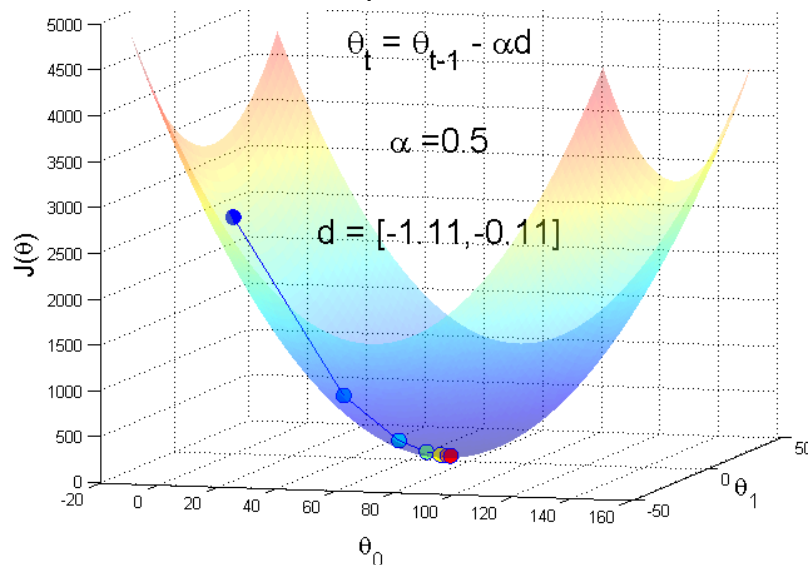
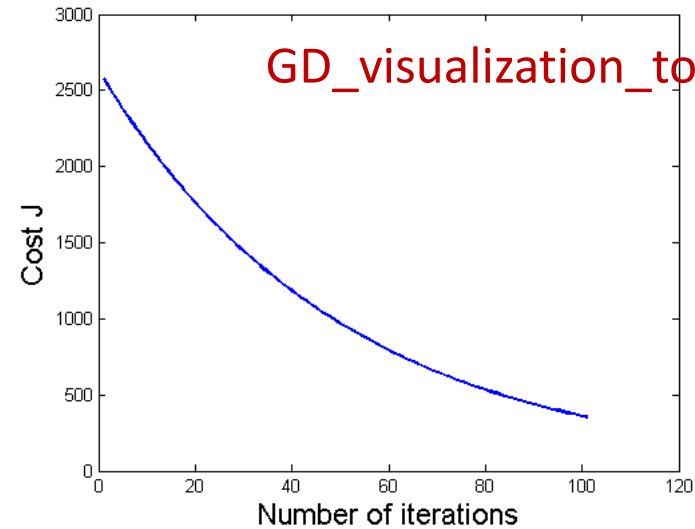
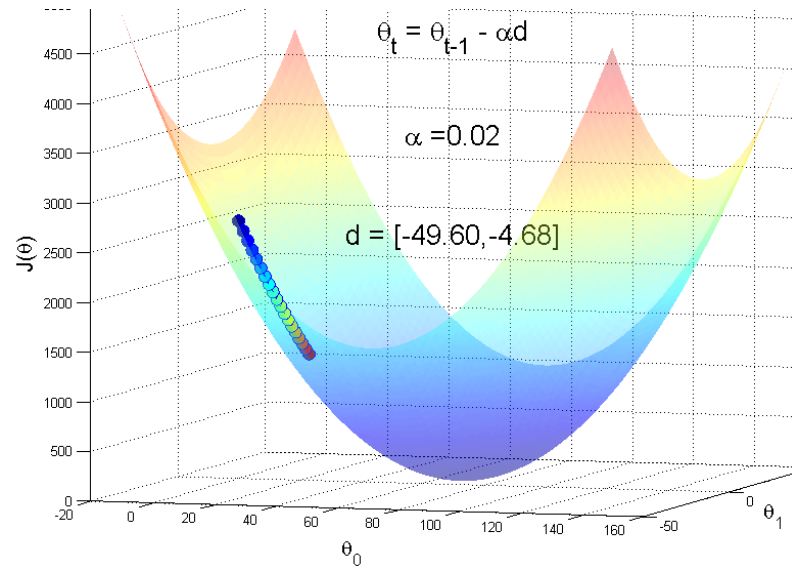
Kako se približavamo minimumu koraci su sve manji jer je gradijent (nagib) sve manji

GD_visualization_slow_convergence.m

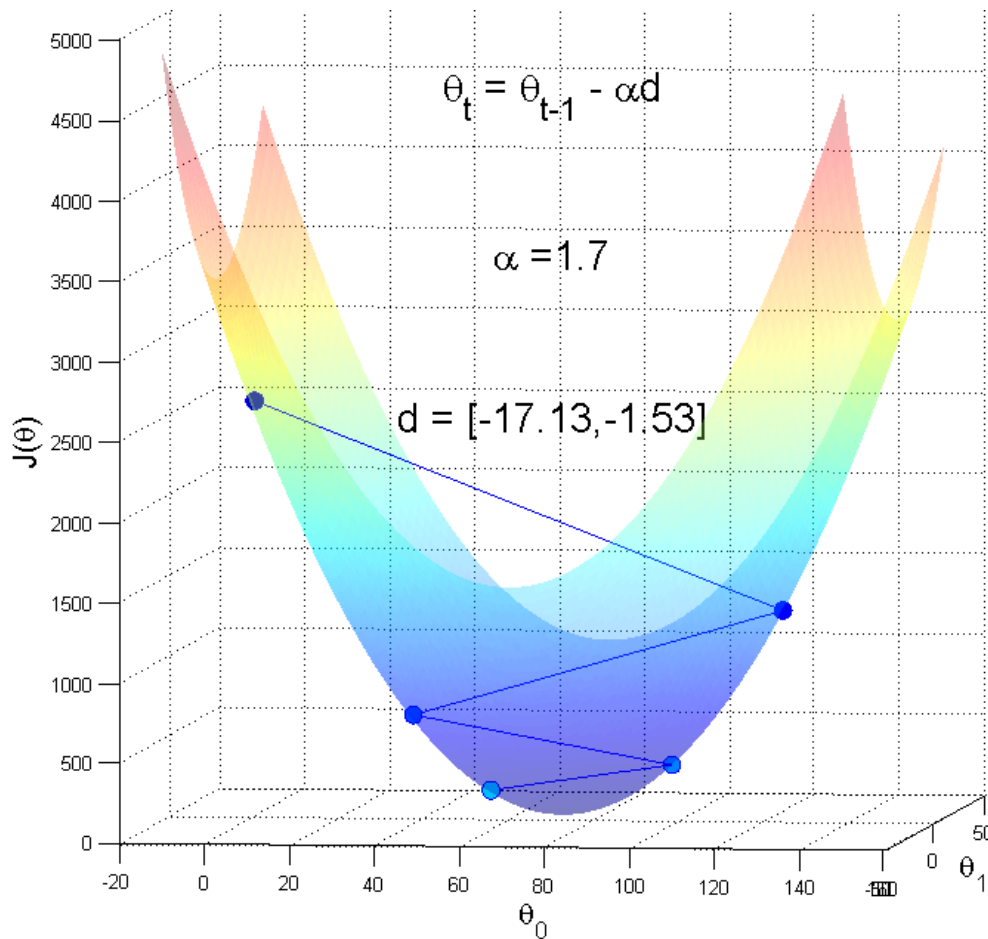
$\alpha = 0.1$

Odabir α : konvergencija

- Za male vrednosti α *gradient descent* je spor



Odabir α : konvergencija



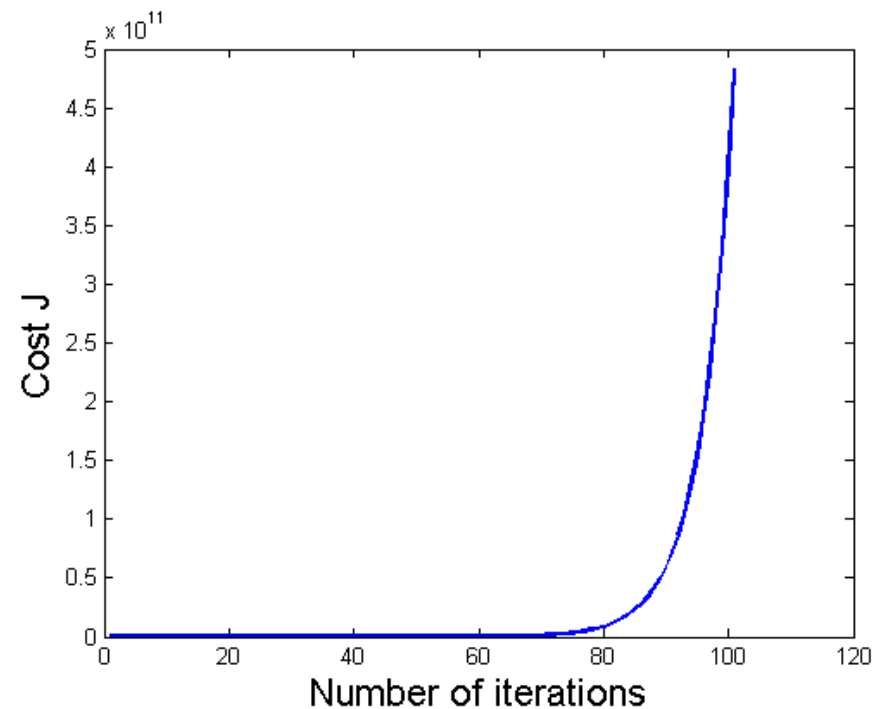
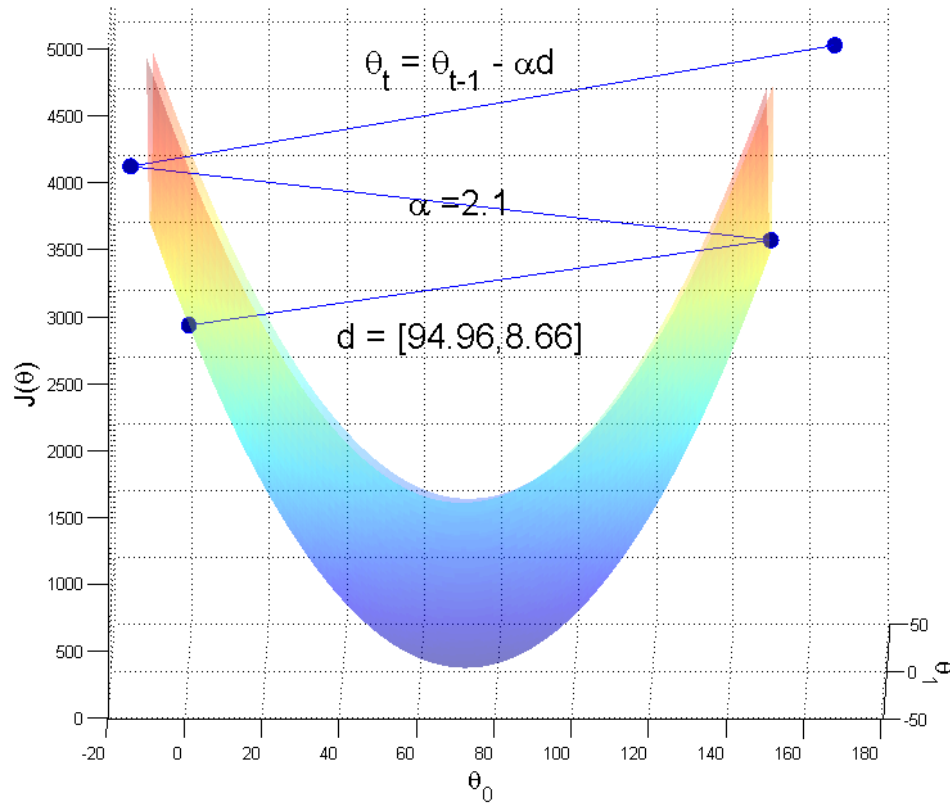
Levo od minimuma vrednost
gradijenta je negativna:
 θ raste

Desno od minimuma
vrednost gradijenta je
pozitivna:
 θ se smanjuje

GD_visualization_left_right.m
 $\alpha = 1.7$

Odabir α : konvergencija

Za preveliko α *gradient descent* može da divergira



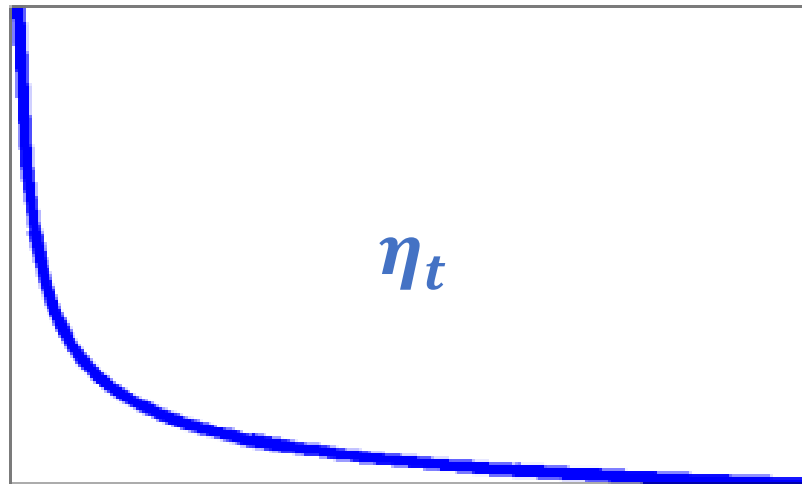
GD_visualization_diverges.m

Odabir α

- Pored fiksnog α , čest izbor je i *stepsize schedule*

- Smanjivanje koraka α sa brojem iteracija:

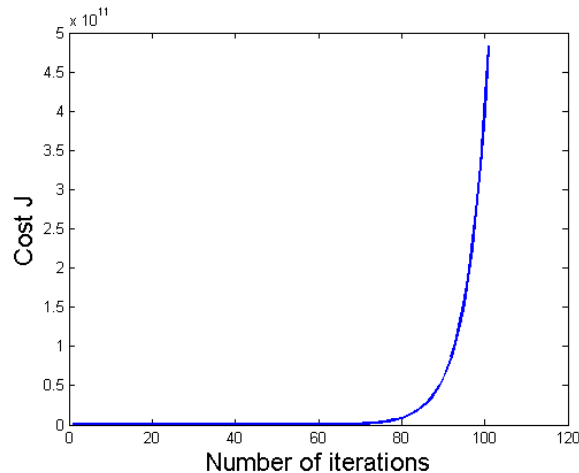
$$\eta_t = \alpha/t; \eta_t = \alpha/\sqrt{t}$$



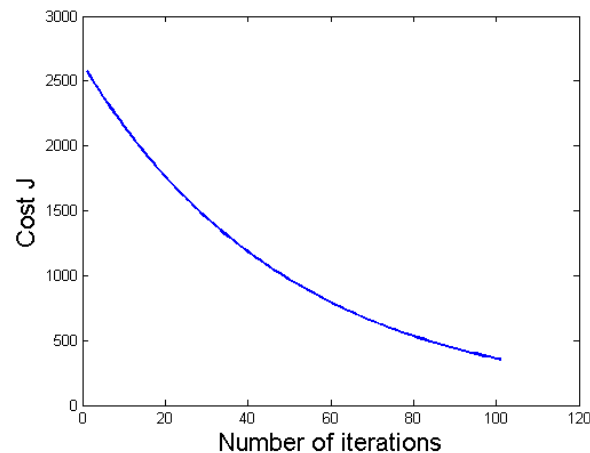
Odabir α : zaključak

Da bismo odredili α , posmatramo grafik:

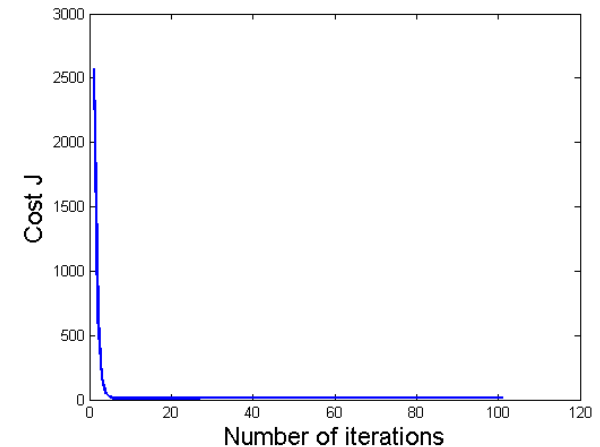
- Y-osa: $J(\theta)$
- X-osa: broj iteracija
- Probati $\alpha = 0.001$, $\alpha = 0.01$, $\alpha = 0.1$, $\alpha = 1, \dots$



Došlo je do divergencije.
Smanjiti α



Algoritam je spor. Nije konvergirao u zadanom broju iteracija. Povećati α



Dobra vrednost α .
Algoritam je brzo konvergirao.

Odabir α : zaključak

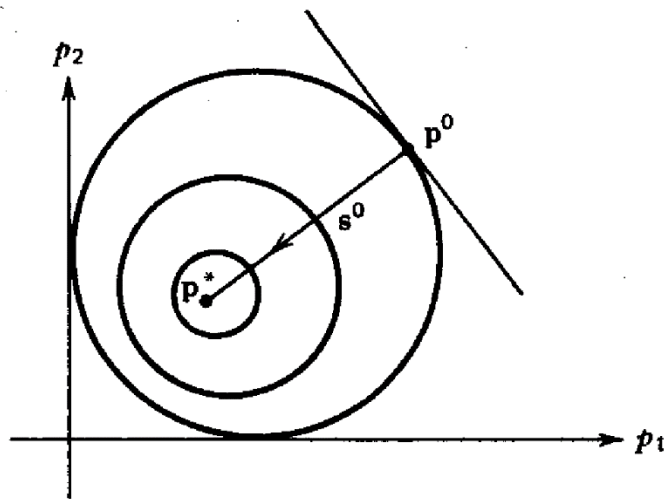
- Alternativa: automatska detekcija konvergencije
- Možemo reći da je algoritam konvergirao kada važi

$$\left\| \frac{\partial}{\partial \theta} J(\theta) \right\| < \varepsilon$$

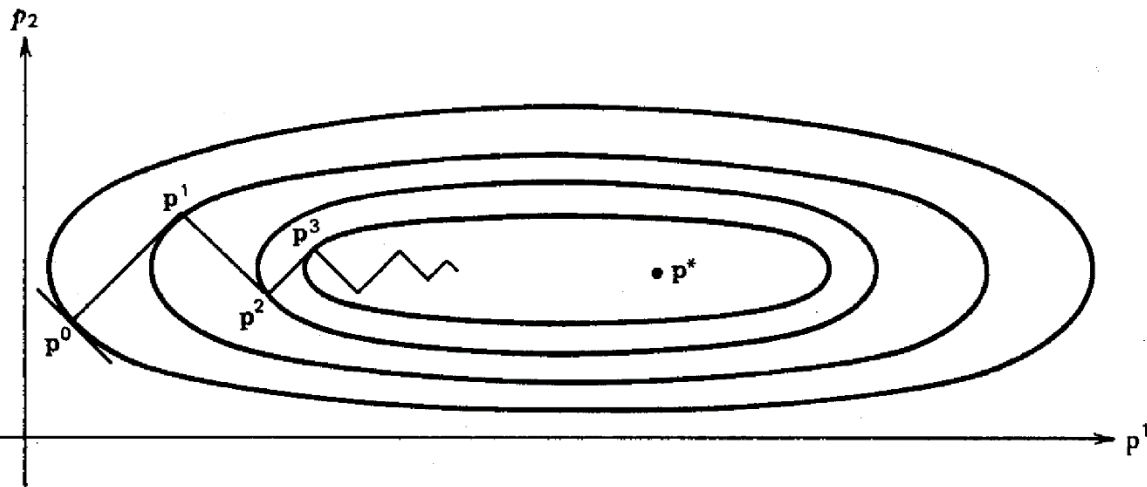
- ε predstavlja prag (malu vrednost, tipa 10^{-5})
- U praksi, šta znači „mala“ vrednost dosta zavisi od podataka

Skaliranje (normalizacija) obeležja

- Vrednosti obeležja treba da budu istog opsega kako bi *gradient descent* brže konvergirao
 - x_1 = SanitationFacilities [0-100 %], x_2 = GDP [597.4 – 143,788.2]



Sa normalizacijom



Bez normalizacije

Skaliranje (normalizacija) obeležja

- **Mean normalization:** skalirati vrednosti obeležja x_d tako da njegova srednja vrednost bude bliska nuli

$$x_d = \frac{x_d - \min(x_d)}{\max(x_d) - \min(x_d)}$$

$$x_d = \frac{x_d - \mu}{\sigma}$$

- μ - srednja vrednost obeležja x_d
- σ - standardna devijacija obeležja x_d

Rezime

- Gradientni spust u praksi
 - početno rešenje
 - α
 - konvergencija
 - značaj skaliranja obeležja
- Sledeće: alternativa gradijentnom spustu za treniranje modela