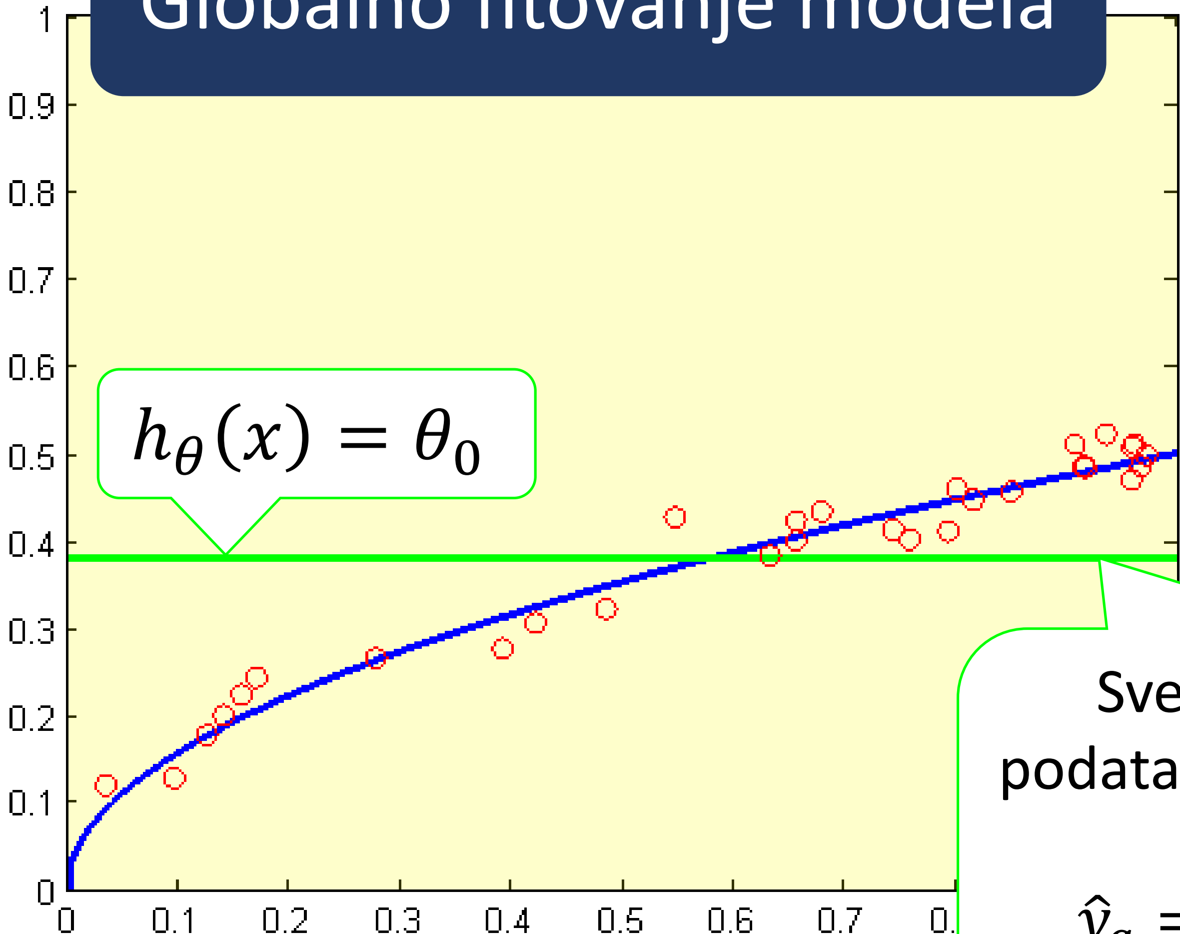
Globalno/lokalno fitovanje modela

Globalno fitovanje modela



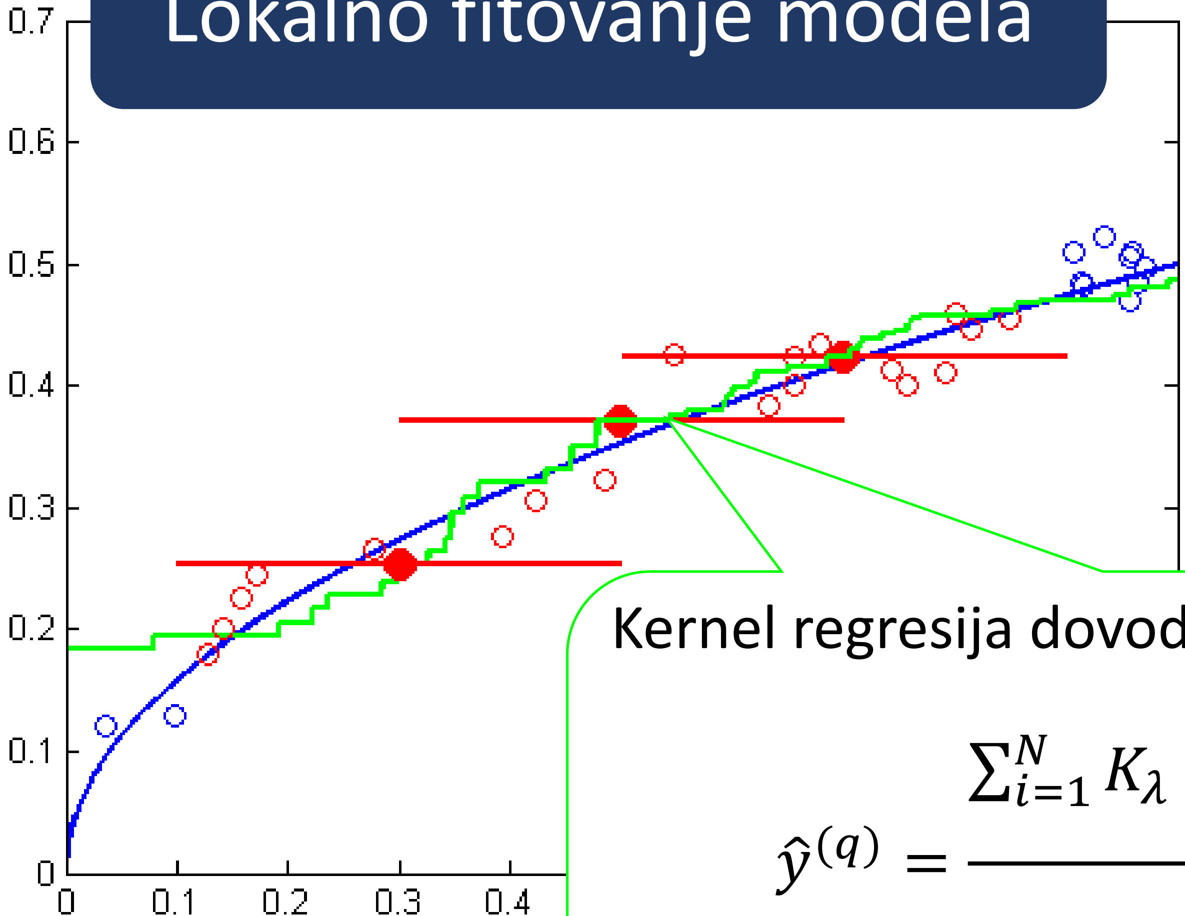
$$h_\theta(x) = \theta_0$$

Sve opservacije iz skupa podataka imaju jednaku težinu:

$$\hat{y}_q = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{\sum_{i=1}^{N} c y_i}{\sum_{i=1}^{N} c}$$

Globalno/lokalno fitovanje modela

Lokalno fitovanje modela



Kernel regresija dovodi do *locally constant fit*

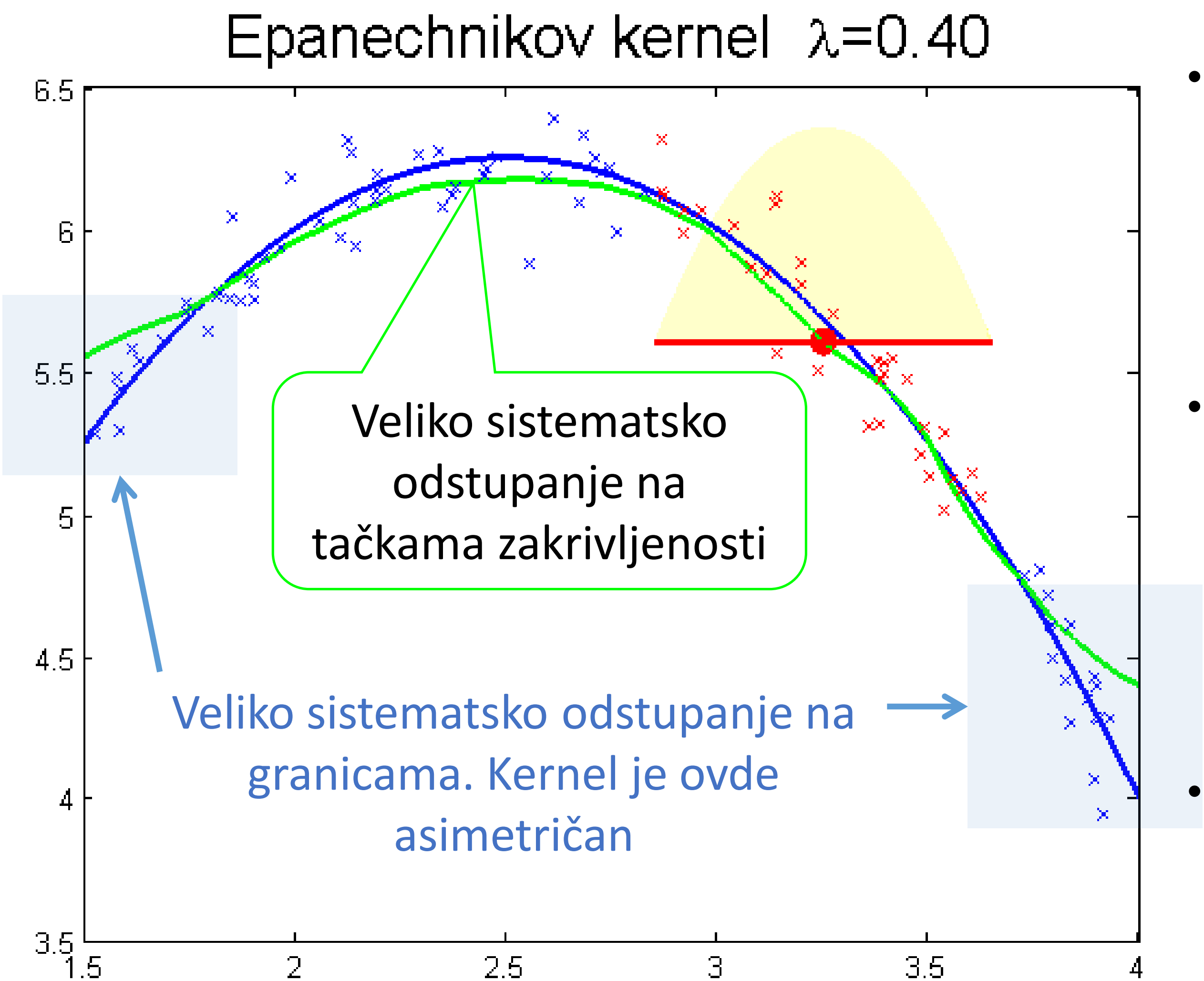
$$\hat{y}^{(q)} = \frac{\sum_{i=1}^{N} K_\lambda\left(\frac{x^{(q)} - x^{(i)}}{\lambda}\right) y^{(i)}}{\sum_{i=1}^{N} K_\lambda\left(\frac{x^{(q)} - x^{(i)}}{\lambda}\right)}$$

Metoda lokalne regresije s težinskim faktorima

- Kernel regresija fituje *konstantnu* funkciju lokalno
  - lokalno → drugačiju za svaku tačku *x*
  - Ovo se naziva *locally weighted averages*

- Da li bismo mogli umesto konstante fitovati liniju ili čak polinomijalnu krivu u svakoj tački?
  - Ovaj pristup se zove Metoda lokalne regresije s težinskim faktorima (*locally weigted regression*)

Metoda lokalne regresije s težinskim faktorima



- Lokalna *linearna* regresija smanjuje sistematsko odstupanje na granicama (sa minimalnim povećanjem varijanse)
- Lokalno fitovana *kvadratna* funkcija ne pomaže na granicama i povećava varijansu, ali pomaže sa sistematskim odstupanjem u tačkama zakrivljenosti
- Preporučeni (podrazumevani) izbor je *lokalna linearna regresija*

Lokalna linearna regresija sa težinskim faktorima

- Linearna regresija:
  - pronaći θ koje minimizuje $\sum_i \left(y^{(i)} - \theta^T x^{(i)}\right)^2$
  - predikcija $\theta^T x$

- Lokalna linearna regresija sa težinskim faktorima:
  - pronaći θ koje minimizuje $\sum_i w^{(i)} \left(y^{(i)} - \theta^T x^{(i)}\right)^2$
  - predikcija $\theta^T x$
  - $w^{(i)}$ su nenegativne vrednosti težina
  - Želimo da damo veću težinu greškama koje su bliže tački za koju određujemo *y*
  - Standardan izbor je $w^{(i)} = \exp\left\{-\frac{\left(x^{(i)} - x\right)^2}{2\lambda^2}\right\}$

Parametarske i neparametarske metode

- Parametarske metode:
  - Linearna regresija (*bez* težinskih faktora)

- Neparametarske (*non-parametric*) metode:
  - metod *k* najbližih suseda
  - kernel regresija
  - lokalna regresija sa težinskim faktorima
  - (i mnoge druge) splines, trees, locally weighted structured regression models...

Parametarske metode

- Fiksni parametari θ (dobijeni na osnovu trening skupa)
- Nakon određivanja θ, za predikciju nam *nije* potreban skup podataka

Neparametarske metode

- Kod *neparametarskih metoda* prilikom predikcije moramo da koristimo trening skup

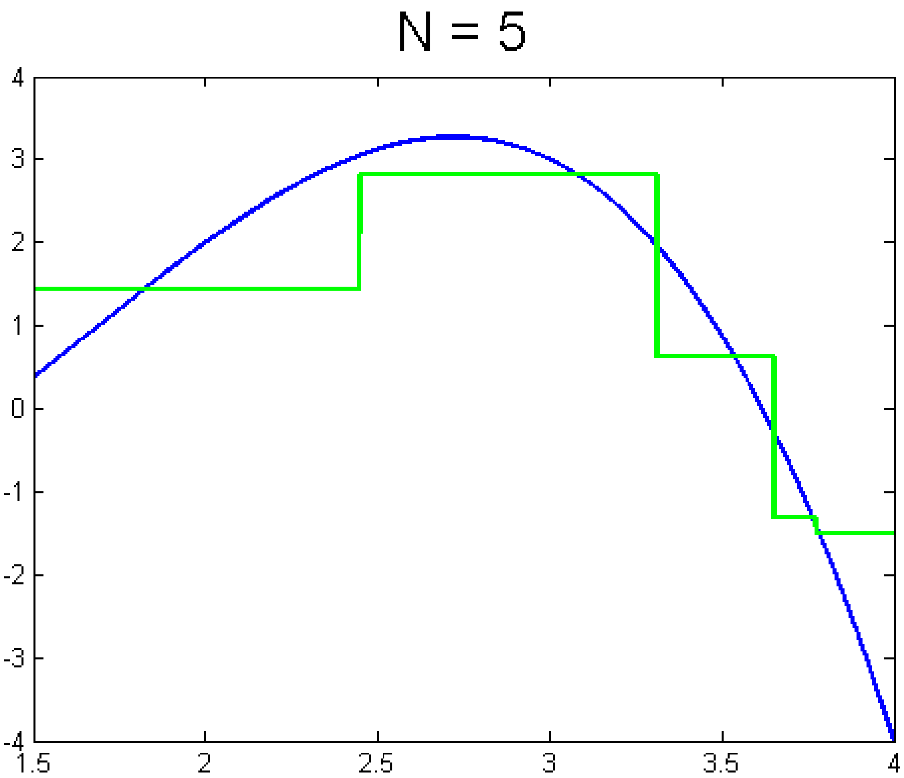
- Količina toga što moramo da čuvamo kako bismo reprezentovali hipotezu *h* raste linearno sa rastom skupa podataka

- Cilj:
  - Fleksibilnost
  - Uvodimo što manje pretpostavki o stvarnoj funkciji *y*
  - Kompleksnost modela raste sa brojem opservacija *N*

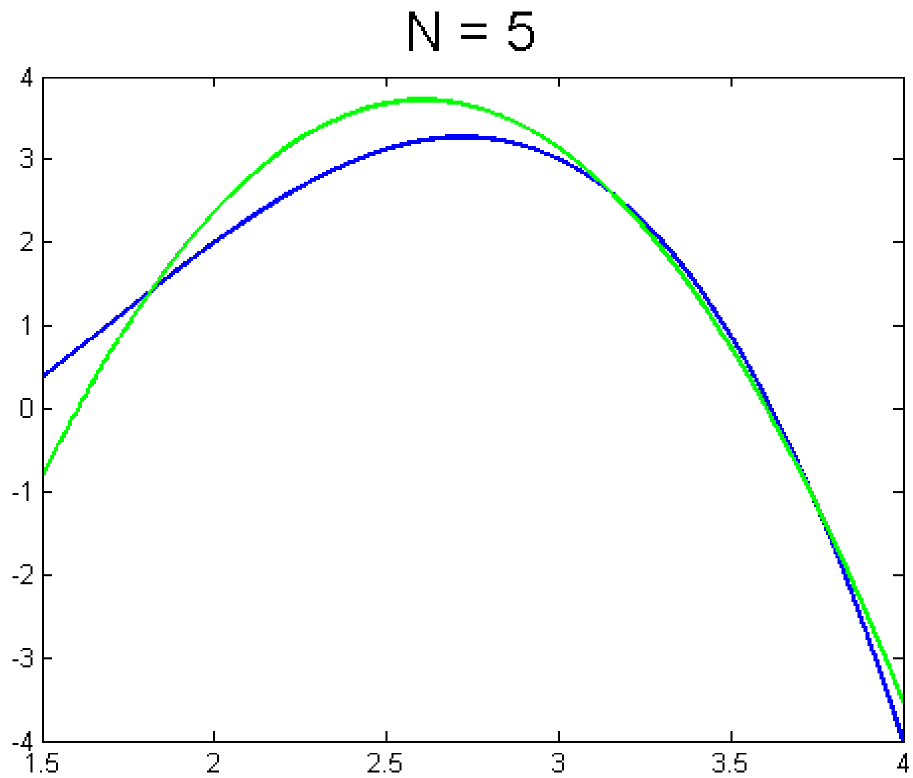
Fleksibilnost

- Ako imamo beskonačnu količinu podataka bez šuma Mean Squared Error 1-NN metode je 0
- Odnosno, sistematsko odstupanje i varijansa su 0
- Ovo ne važi za parametarske modele – uvek ima nešto sistematskog odstupanja
- Razmotrimo ovo na primeru:
  - stvarna funkcija je 3. stepena
  - fitujemo 1-NN i polinom 2. stepena

Granično ponašanje *k*-NN (bez šuma)

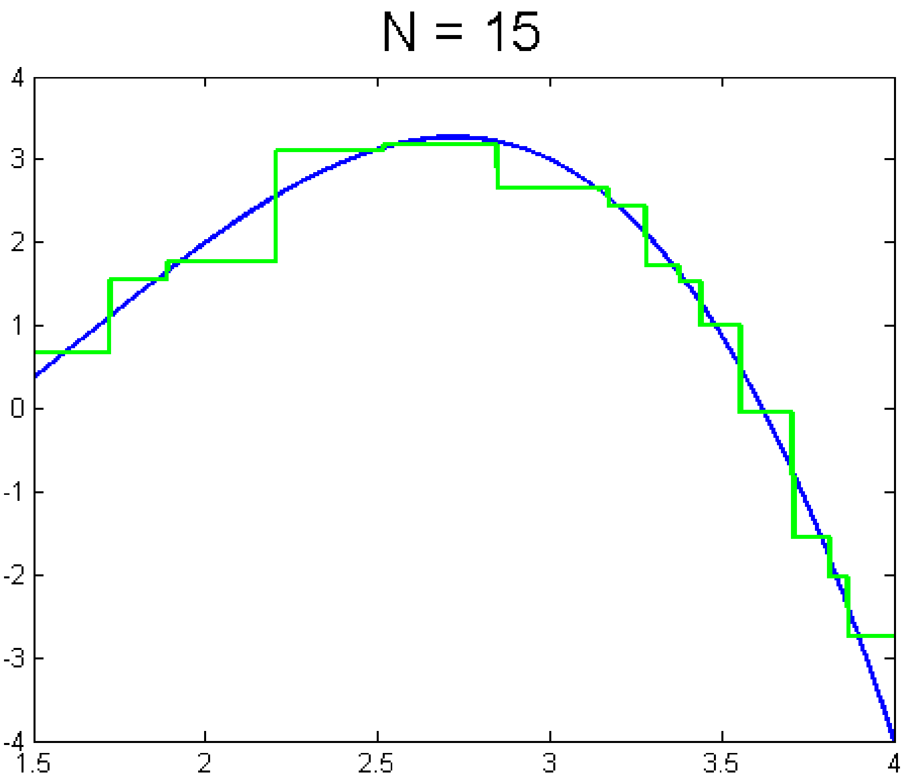


1-NN

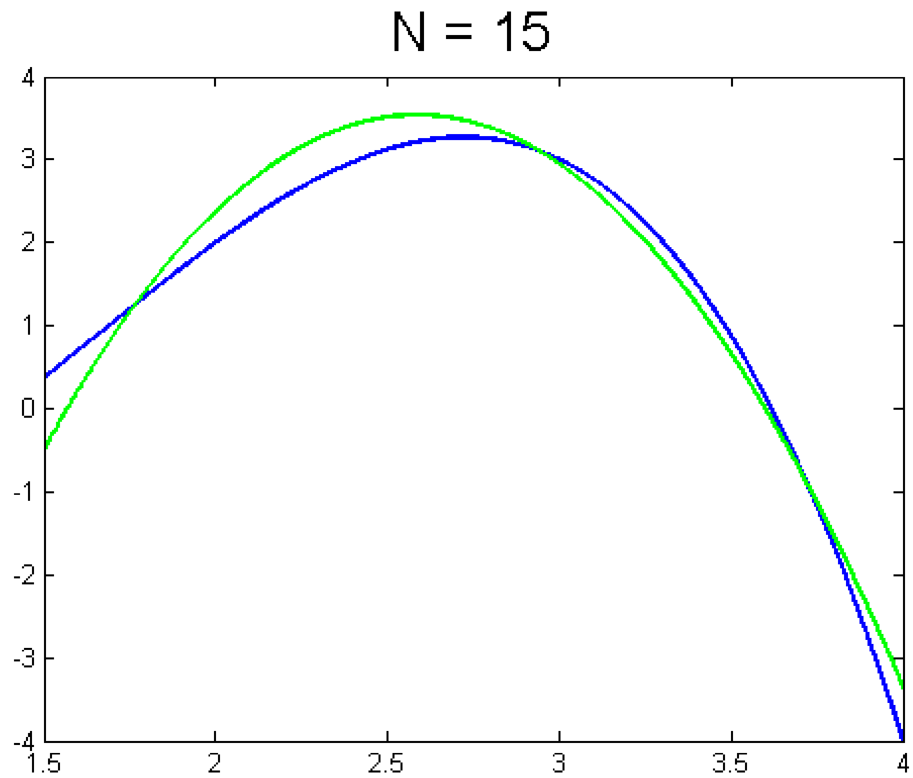


$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$

Granično ponašanje *k*-NN (bez šuma)

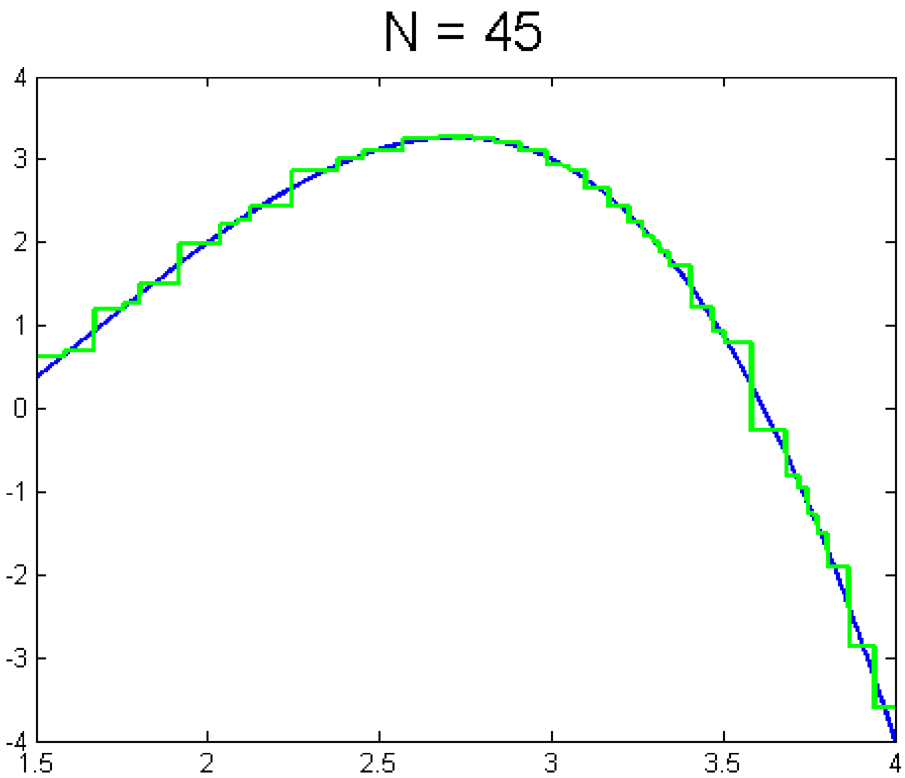


1-NN

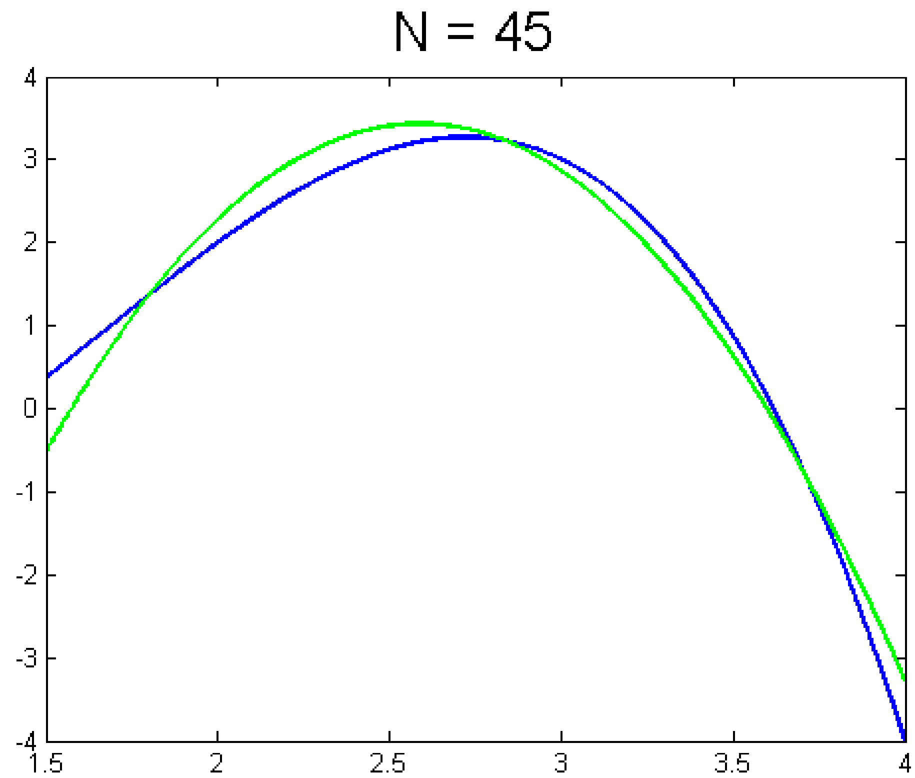


$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

Granično ponašanje *k*-NN (bez šuma)

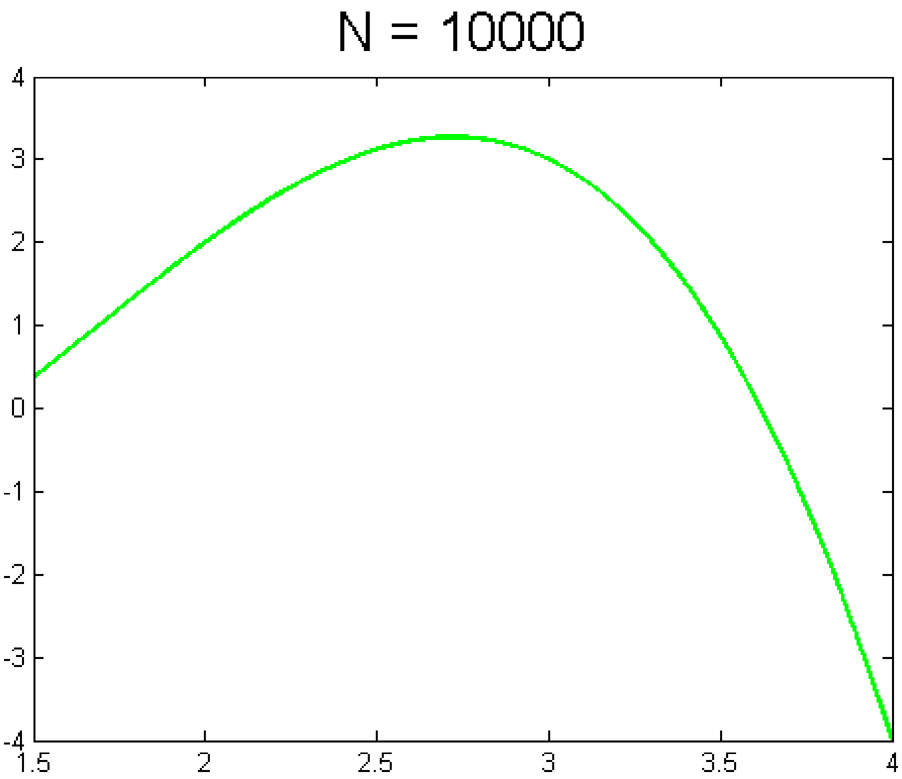


1-NN

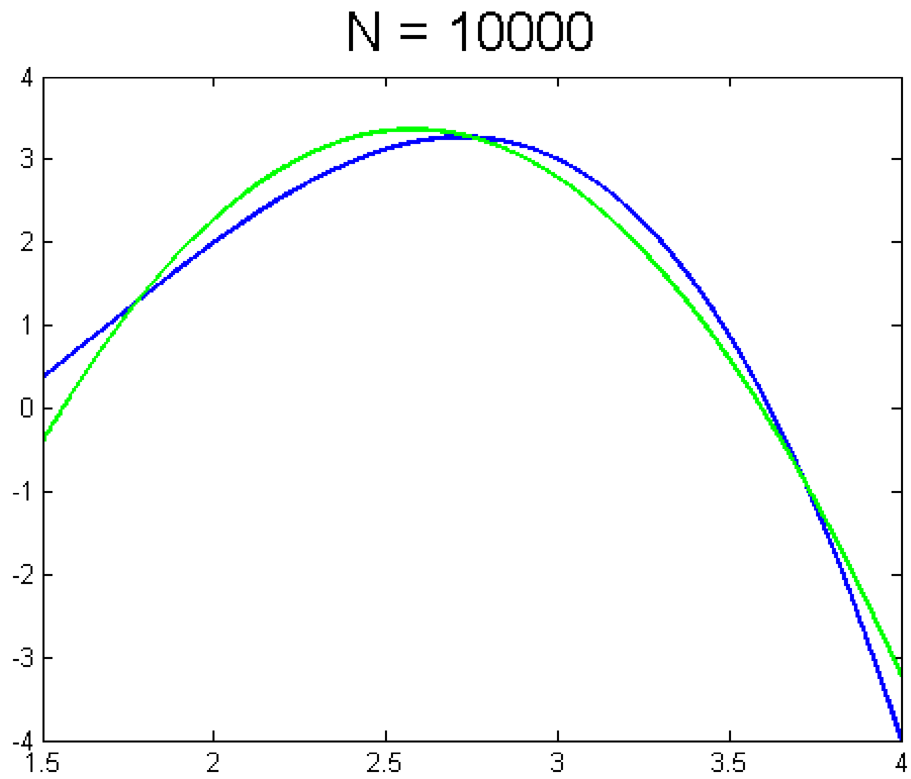


$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$

Granično ponašanje *k*-NN (bez šuma)

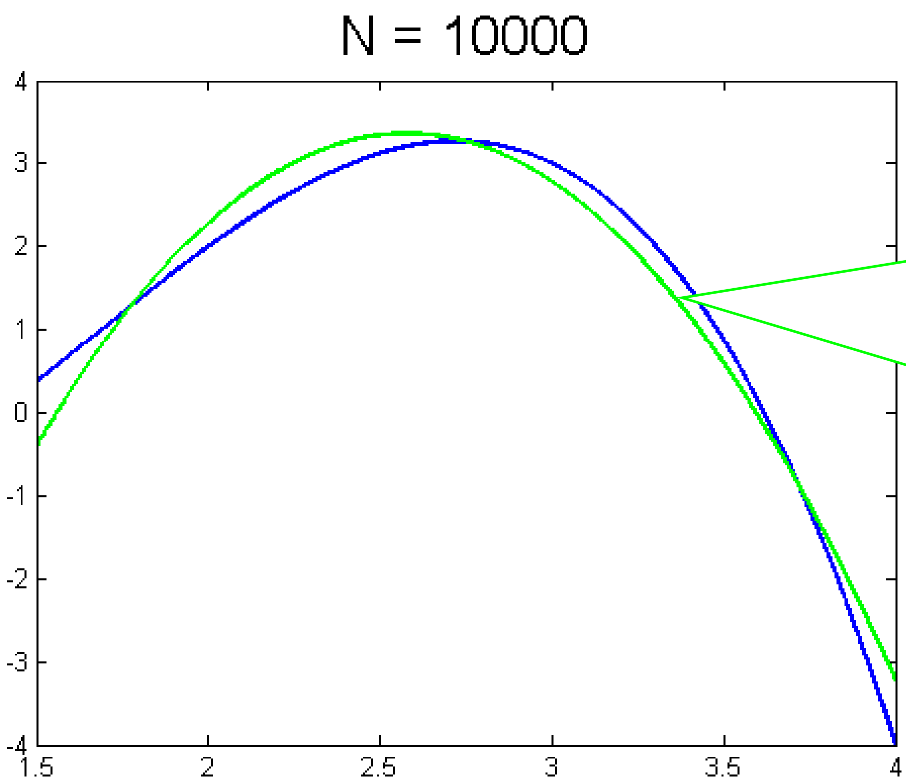


1-NN



$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$

Zaključak – parametarski metod



$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

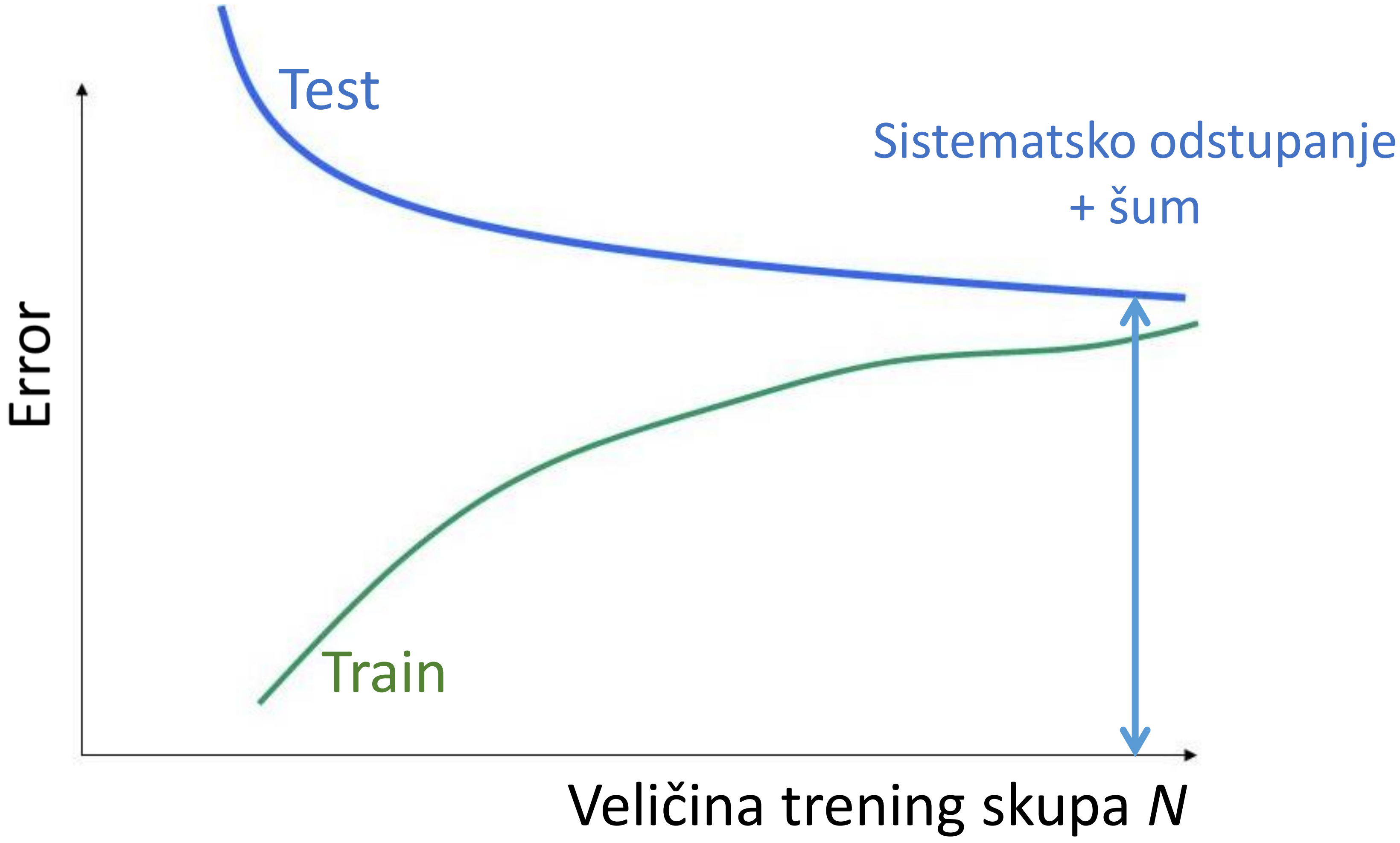
Parametarske metode:

čak i za $N = \infty$ i $\varepsilon = 0$, generalizaciona greška neće spasti na 0 zbog sistematskog odstupanja

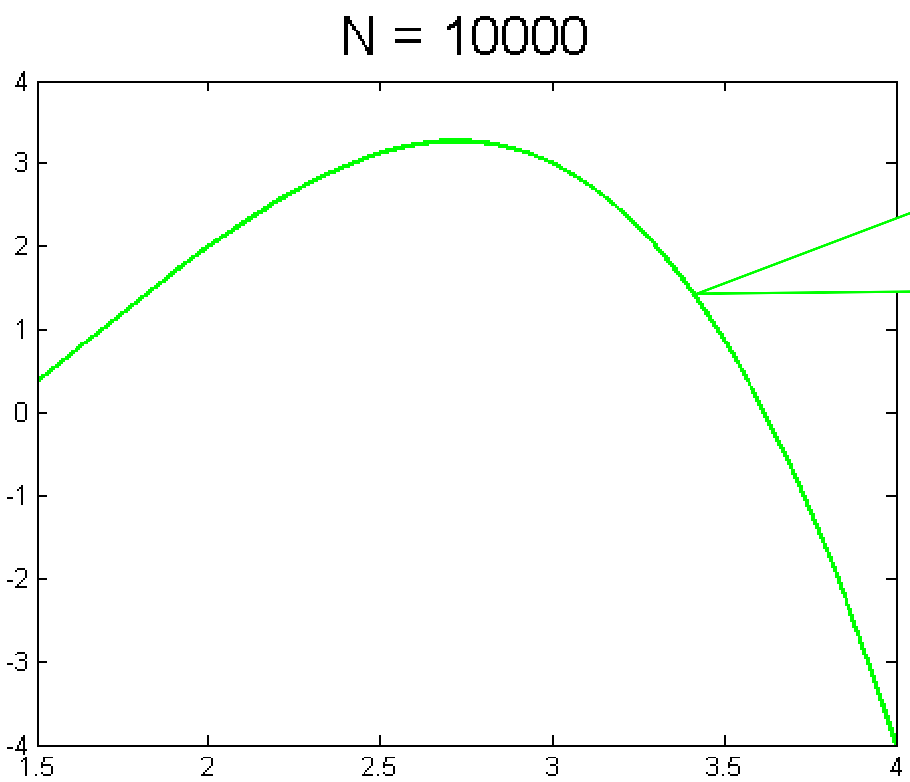
(osim ako smo pogodili model, tj., ciljna funkcija ima identičan oblik kao model)

Zaključak – parametarski metod

(Fiksirana je kompleksnost modela)



Zaključak – neparametarski metod

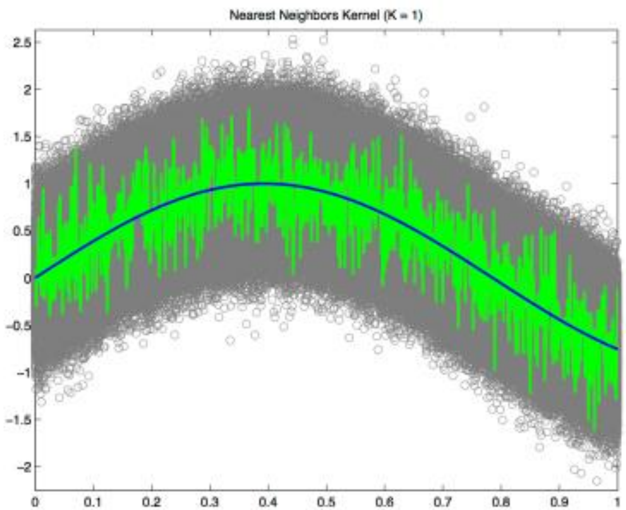


1-NN

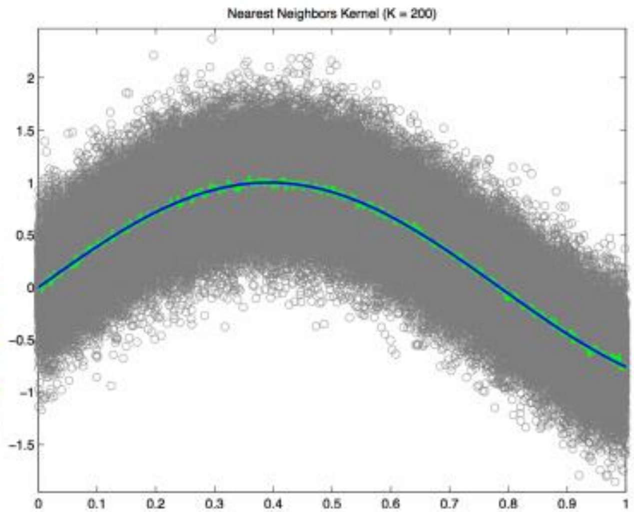
Neparametarske metode:

za $N = \infty$ i $\varepsilon = 0$, generalizaciona greška će spasti na 0

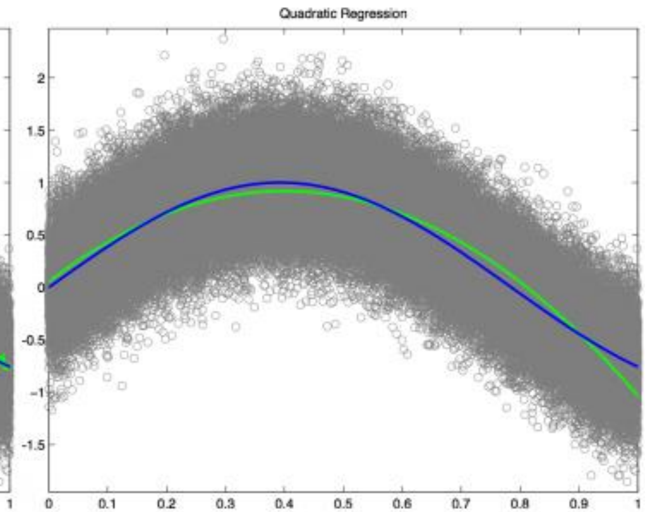
Granično ponašanje *k*-NN (sa šumom)



1-NN fit



200-NN fit



Quadratic fit

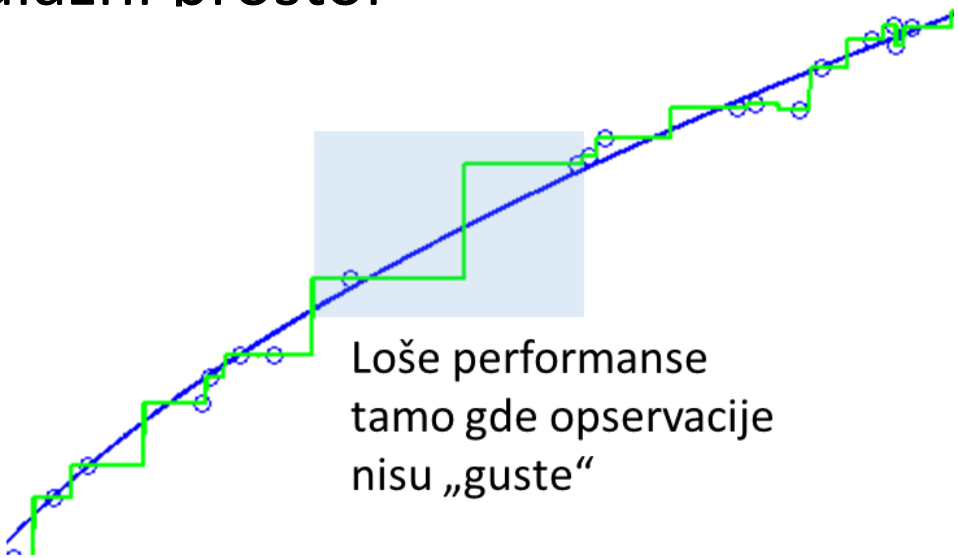
Ako držimo *k* fiksirano, učimo samo lokalno (prilagođavamo se šumu)

MSE k-NN metode će spasti na 0 ako dozvolimo da *k* raste sa N

Kod parametarskih metoda je i u ovom slučaju prisutno sistematsko odstupanje

Mane neparametarskih modela

- *k*-NN i kernel regresija rade dobro kada opservacije „pokrivaju“ ulazni prostor



- Ovo je dobro za dobro za malo D ($D \leq 4$) i veliko N
- Imamo dovoljno tačaka da u svakom susedstvu možemo izračunati prosek

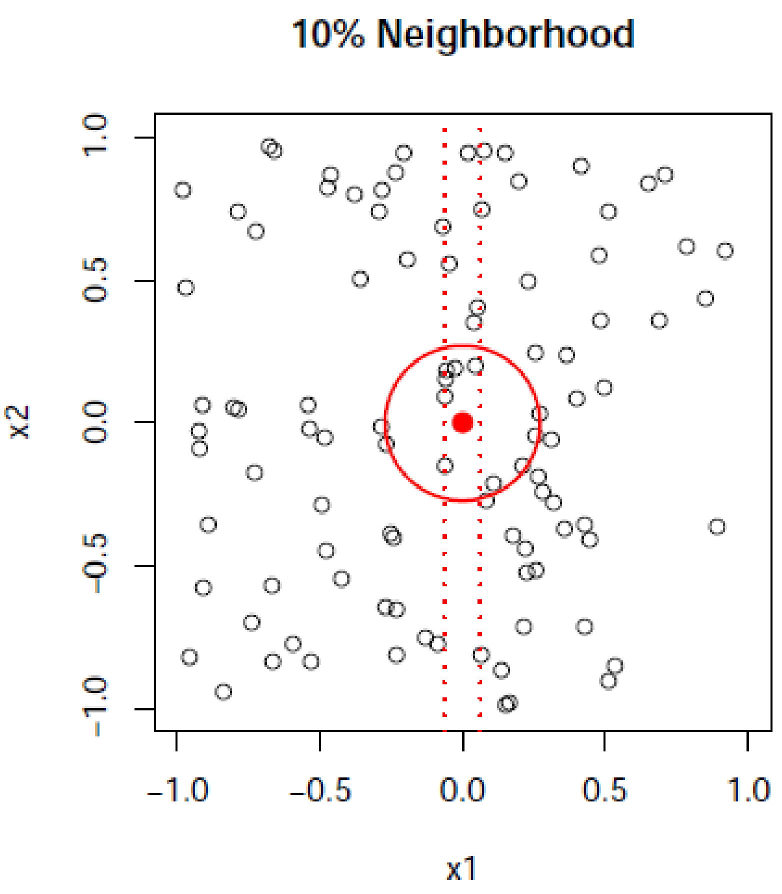
Mane neparametarskih modela

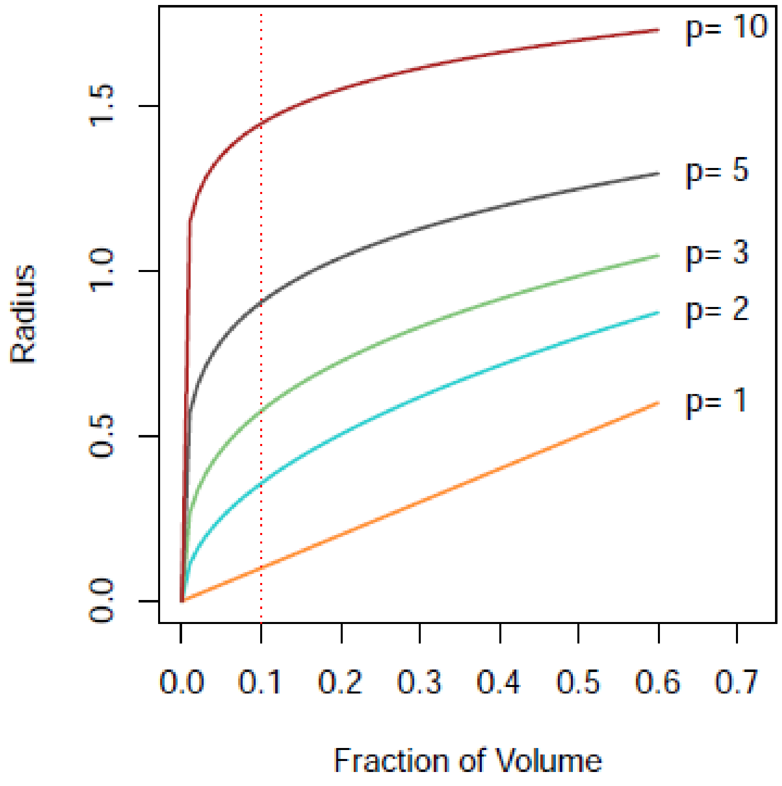
- Međutim, ove metode mogu imati veoma loše performanse za veliko D

- Razlog: prokletstvo dimenzionalnosti – najbliži susedi mogu biti veoma daleko kada imamo visok broj dimenzija

Prokletstvo dimenzionalnosti

Primer: želimo da napravimo susedstvo koje sadrži 10% primera od *N* dostupnih





Kako se borimo sa prokletstvom dimenzionalnosti?

- Pridodajemo strukturu našem modelu
- Najjednostavnija struktura je linearan model. Linearan model je važan predstavnik parametarskih (struktuiranih) modela:
$$h_L(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_D x_D$$
- Specificiran je sa $D + 1$ parametrom
- Ovaj model ne pati od prokletstva dimenzionalnosti jer se ne oslanja ni na kakva lokalna svojstva (uprosečavanja najbližih suseda)
- Gotovo nikada nije tačan, ali služi kao dobra i interpretabilna aproksimacija prave (nepoznate) funkcije y

Sumarizacija

Neparametarski modeli

Slabe pretpostavke
(mogu da modeluju izuzetno
kompleksne sisteme)

Prokletstvo dimenzionalnosti
je veliki problem

Potraga za najbližim
susedstvom je vremenski
zahtevna, pogotovo za veliko
N i D

Parametarski modeli

Nameću snažne pretpostavke
o ciljnoj funkciji
(koje ne moraju da važe)

Prokletstvo dimenzionalnosti
je manji problem