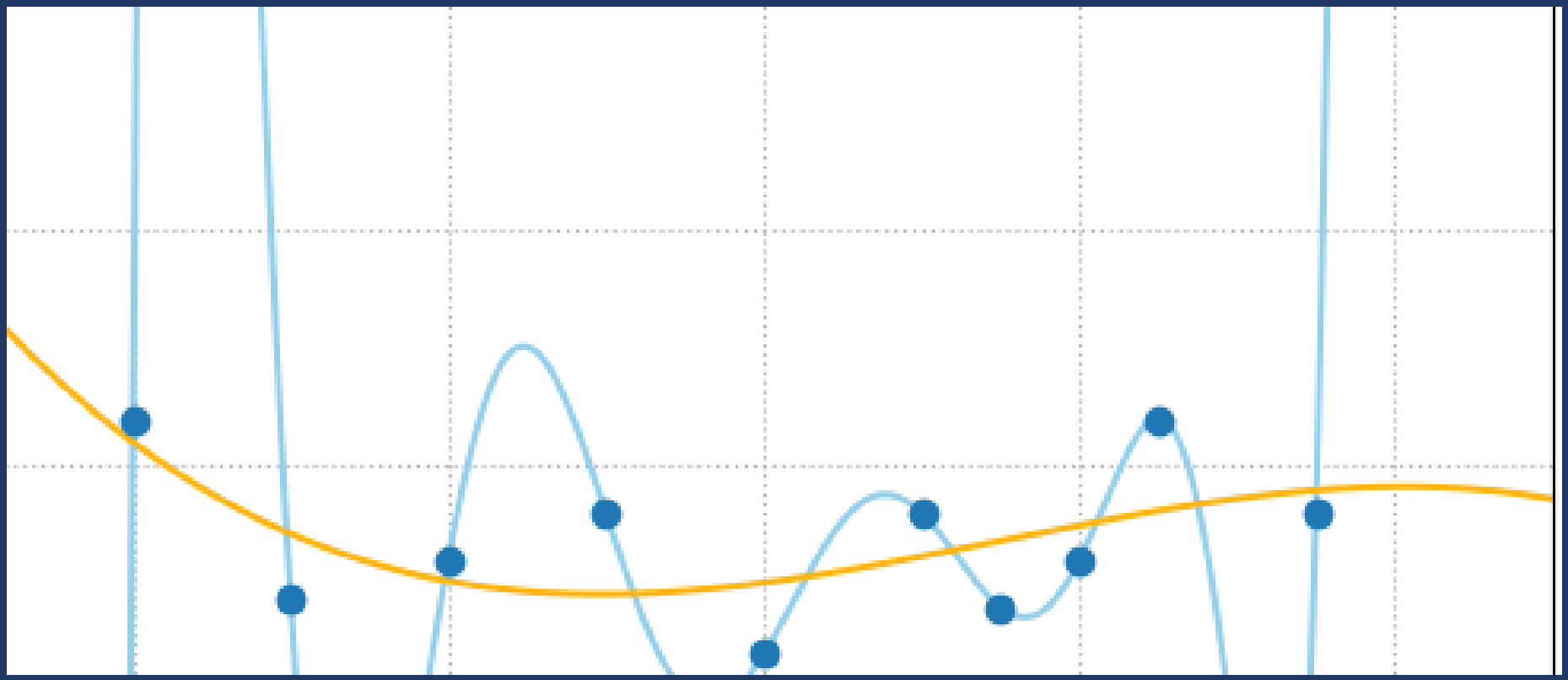


# Regularizacija

Ridge (L2), Lasso (L1), Elastic Net

# Ordinary Least Squares (OLS)

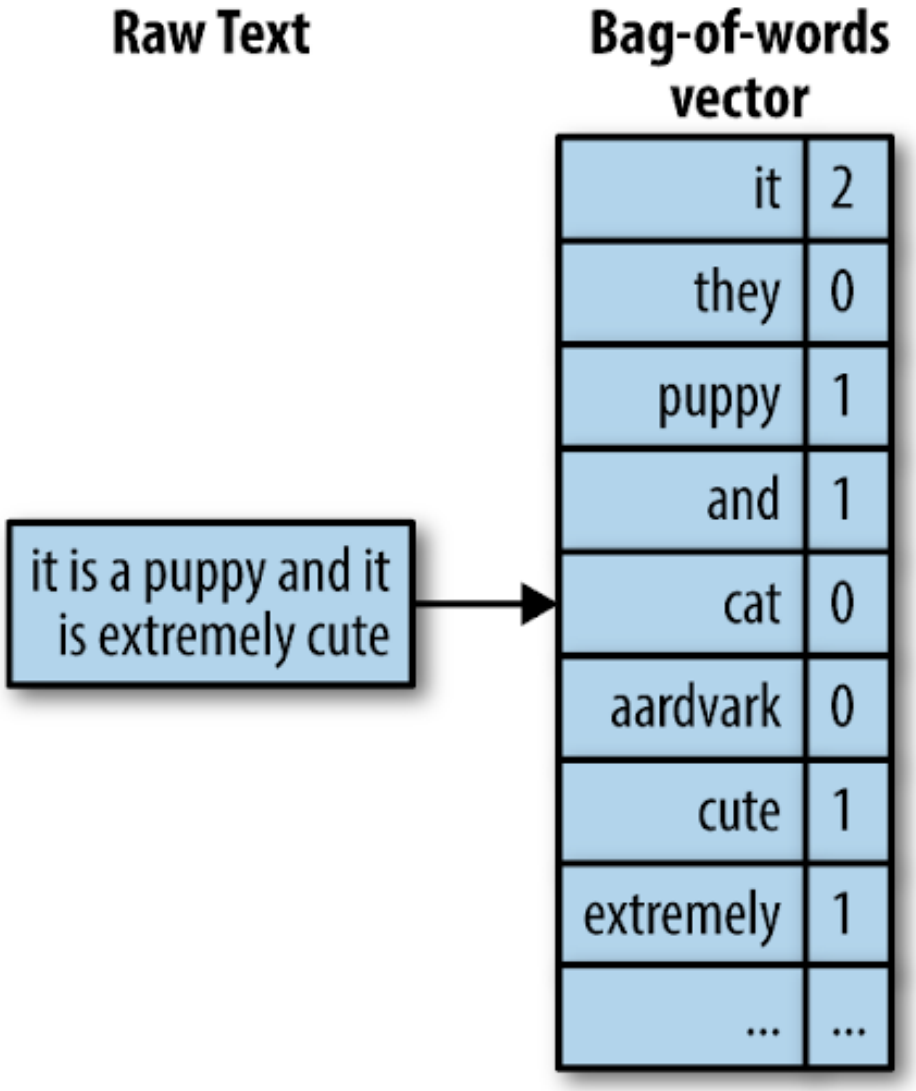
- OLS model:

$$\theta = (X^T X)^{-1} X^T \mathrm{y}$$

- Problem: $X^T X$ može biti singularna matrica
  - Kada je broj obeležja veći od broja instanci $(D > N)$
  - Kada su prediktori multikolinearni

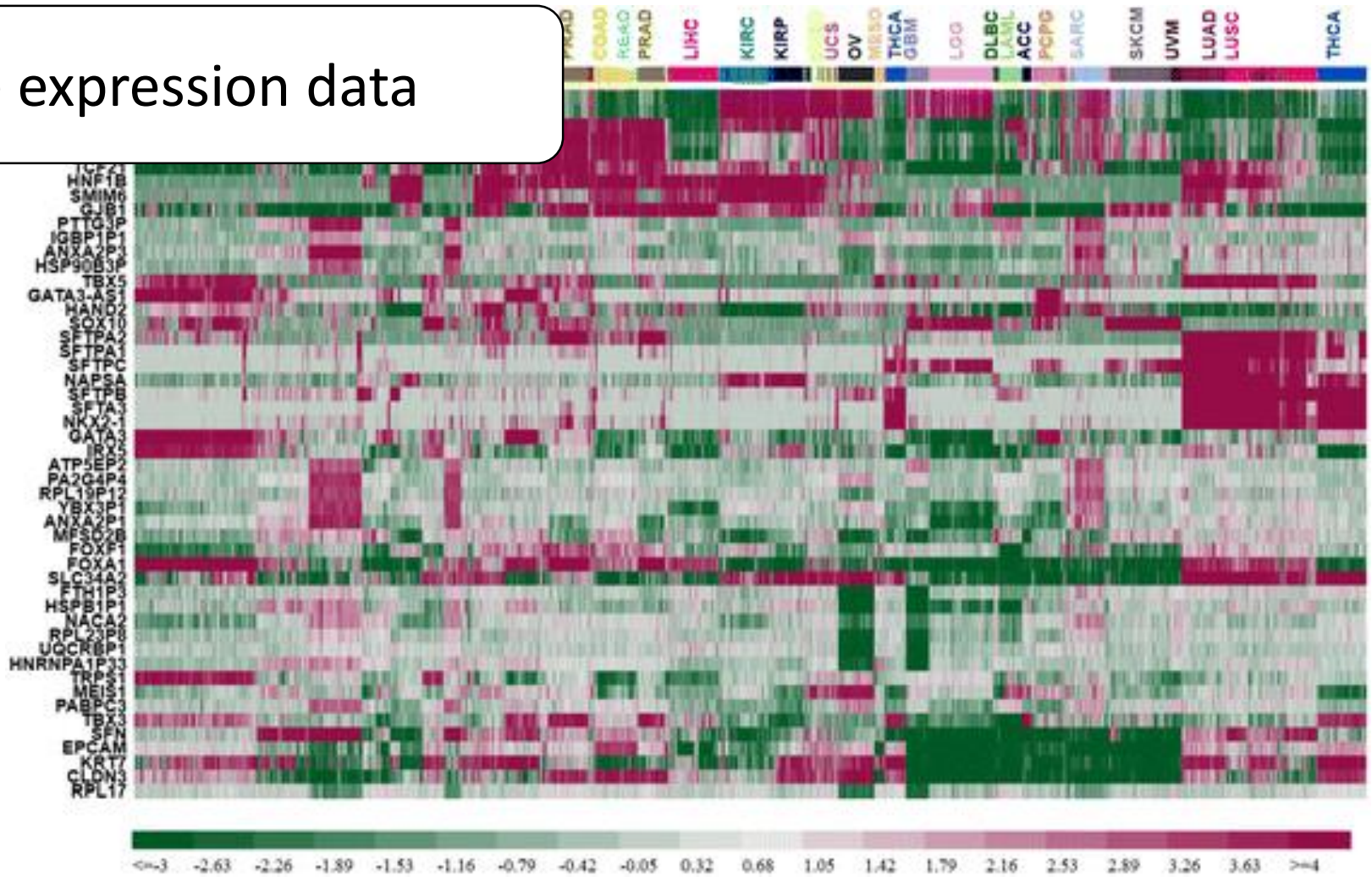
# Slučaj $D > N$ nije redak u praksi

Klasifikacija tekstualnih dokumenata



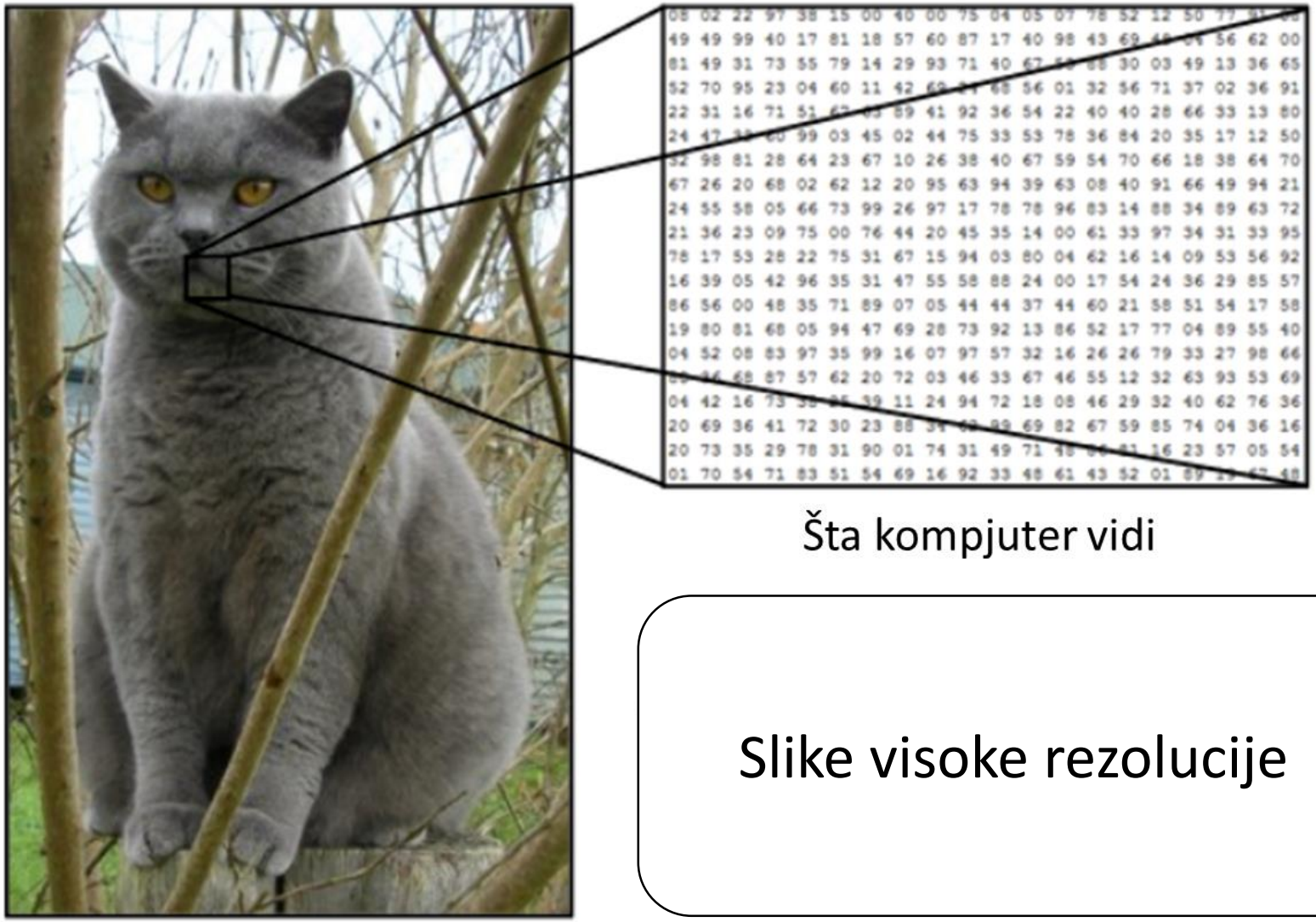
# Slučaj $D > N$ nije redak u praksi

Gene expression data




Li, Y., Kang, K., Krahn, J.M., Croutwater, N., Lee, K., Umbach, D.M. and Li, L., 2017. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC genomics*, *18*(1), p.508.

# Slučaj $D > N$ nije redak u praksi



Šta kompjuter vidi

Slike visoke rezolucije

# Multikolinearnost

- Jedno od obeležja predstavlja linearnu kombinaciju drugih obeležja:

$$c_1 f_1 + c_2 f_2 + \cdots + c_k f_k = f_j$$

- Perfektna multikolinearnost
  - $X^T X$ nema rešenje
  - Ali se retko dešava. Obično je prisutan šum i varijable su samo snažno korelirane
  - Međutim, i ovo je problem

# Snažna korelacija je problematična

- Obeležja

$$x_1 \text{ i } x_2, \text{ gde je } x_1 = x_2$$

- Stvarna (nepoznata) ciljna funkcija

$$y = x_1$$

- Dva alternativna modela:

  Model 1: $y = 0.5 \cdot x_1 + 0.5 \cdot x_2$

  Model 2: $y = 1000 \cdot x_1 - 999 \cdot x_2$

- Neka je za neku instancu $x_2 = 0.95 \cdot x_1$ (šum)
  - Model 1: $y = 0.5 \cdot x_1 + 0.5 \cdot 0.95 \cdot x_1 = 0.975 \cdot x_1$
  - Model 2: $y = 1000 \cdot x_1 - 999 \cdot 0.95 \cdot x_1 = 50.95 \cdot x_1$

# Multkolinearnost nije retka u praksi

- Reči u tekstu se ne javljaju u potpunosti nezavisno jedna od druge

- Ekspresije mnogih gena su u velikoj korelaciji

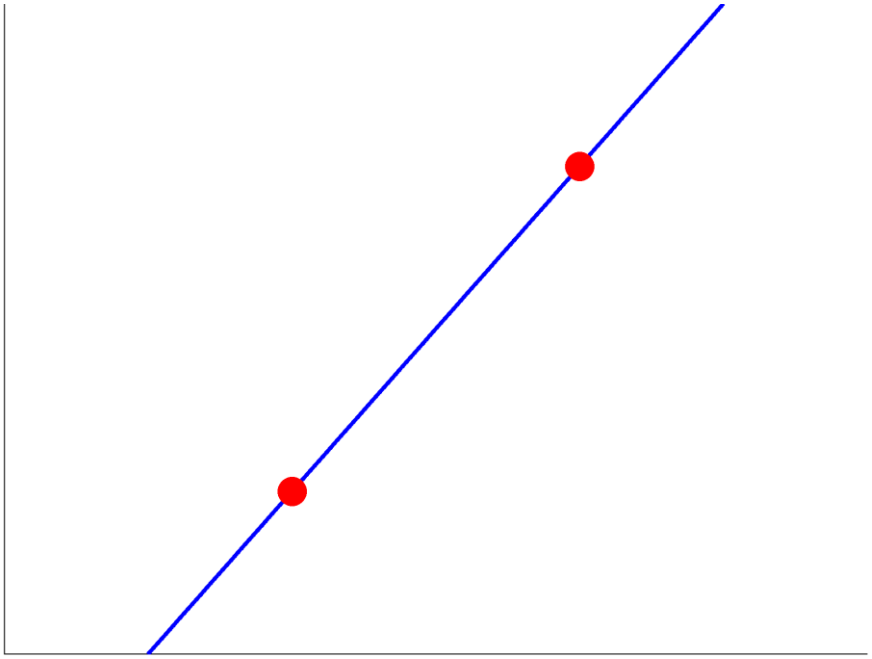
# Ordinary Least Squares (OLS)

- OLS model:

$$\theta = (X^T X)^{-1} X^T \mathrm{y}$$

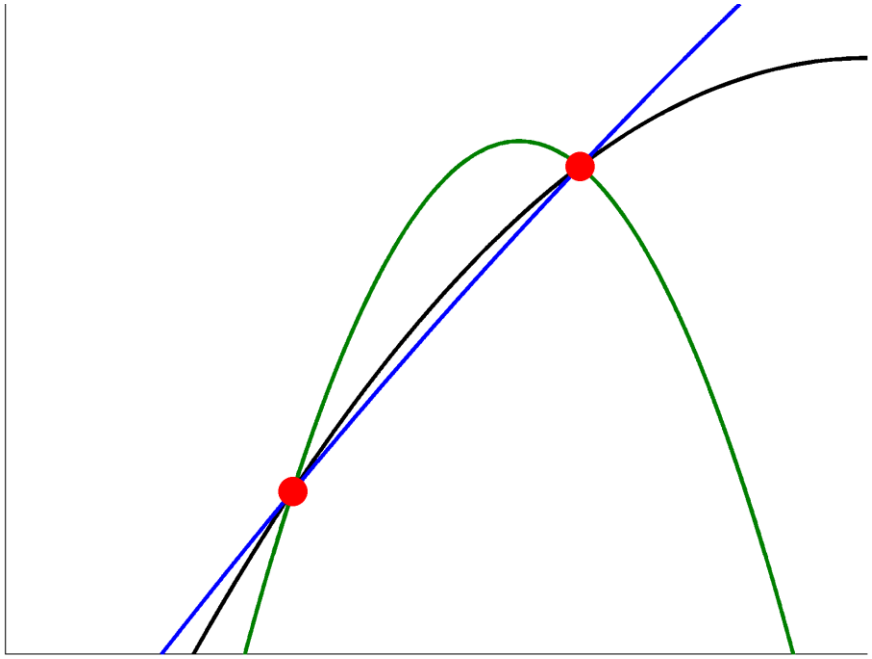
- $X^T X$ je singularna matrica znači da je sistem je neodređen
  - Imamo beskonačno mnogo rešenja i ne znamo koje da odaberemo

# Neodređen sistem – primer



Određen sistem: jedno obeležje

- Linija koja se savršeno uklapa u podatke

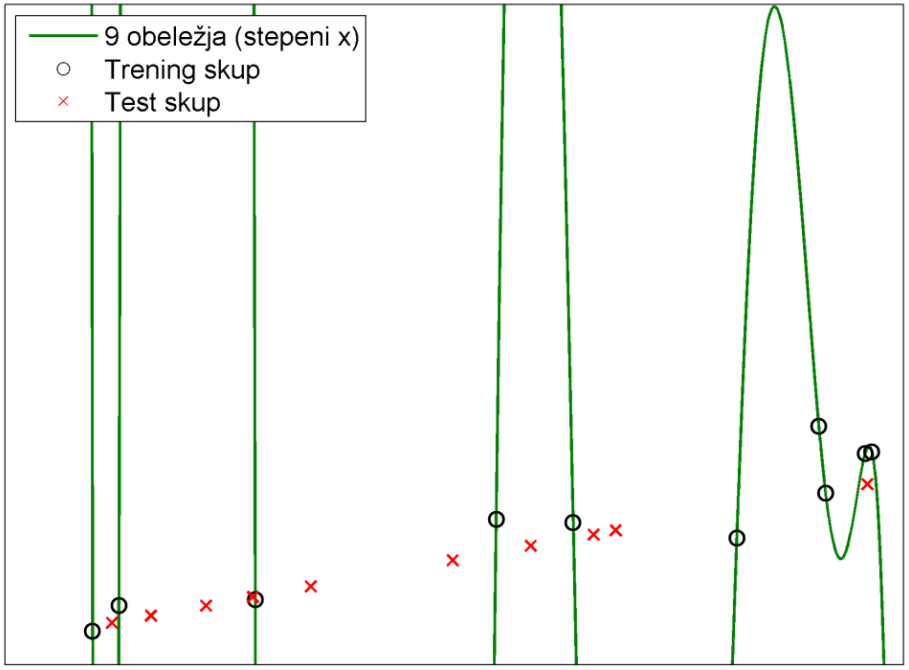


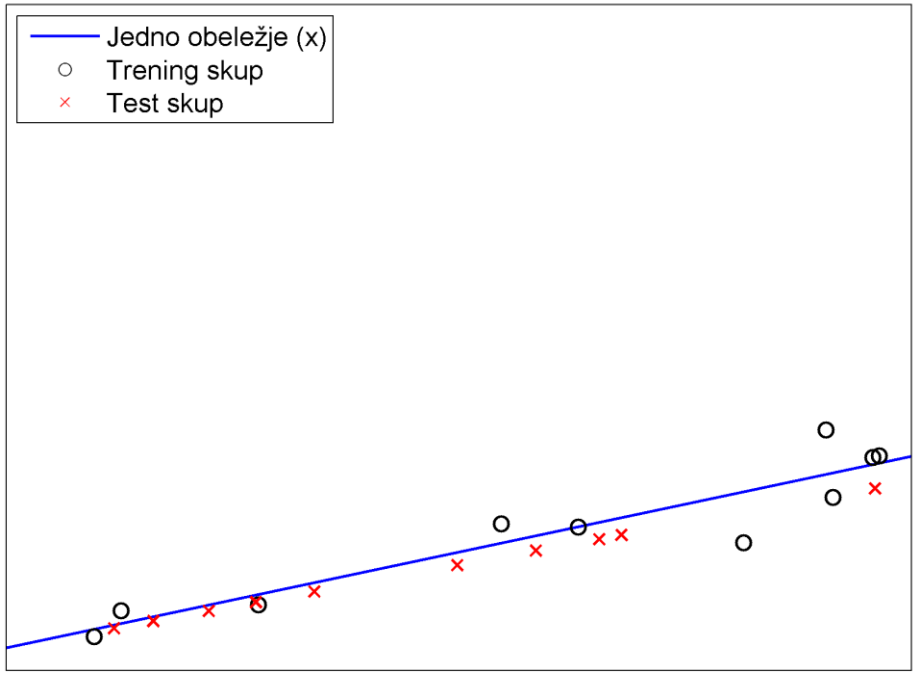
Neodređen sistem: više obeležja

- Svaki model se savršeno uklapa u podatke

# Mnogo varijabli – preprilagođavanje

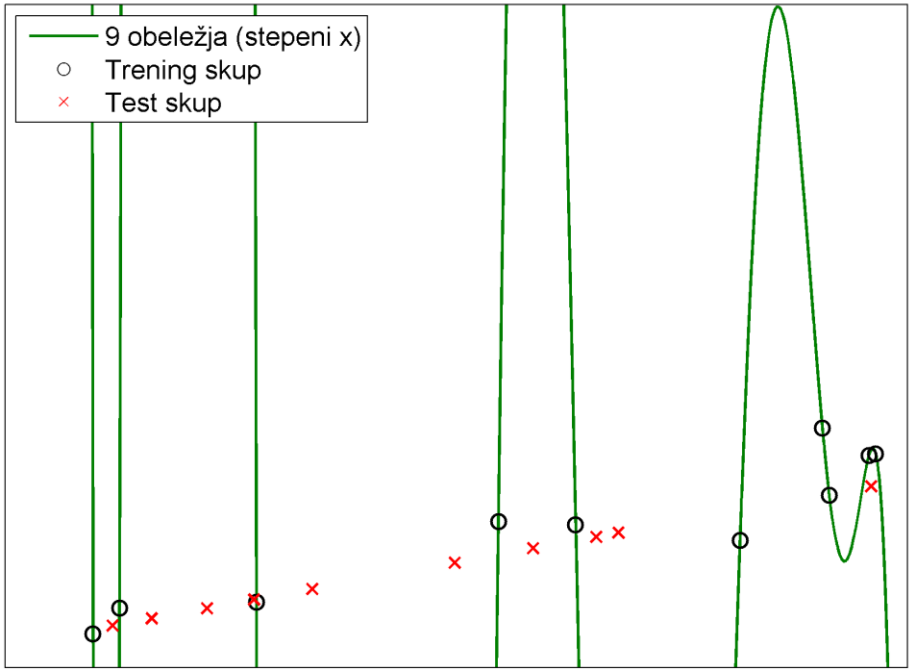
- Za dobru generalizaciju je ključno upravljanje prilagodljivošću modela
- Dobra prilagođenost modela trening podacima ne obezbeđuje dobru generalizaciju

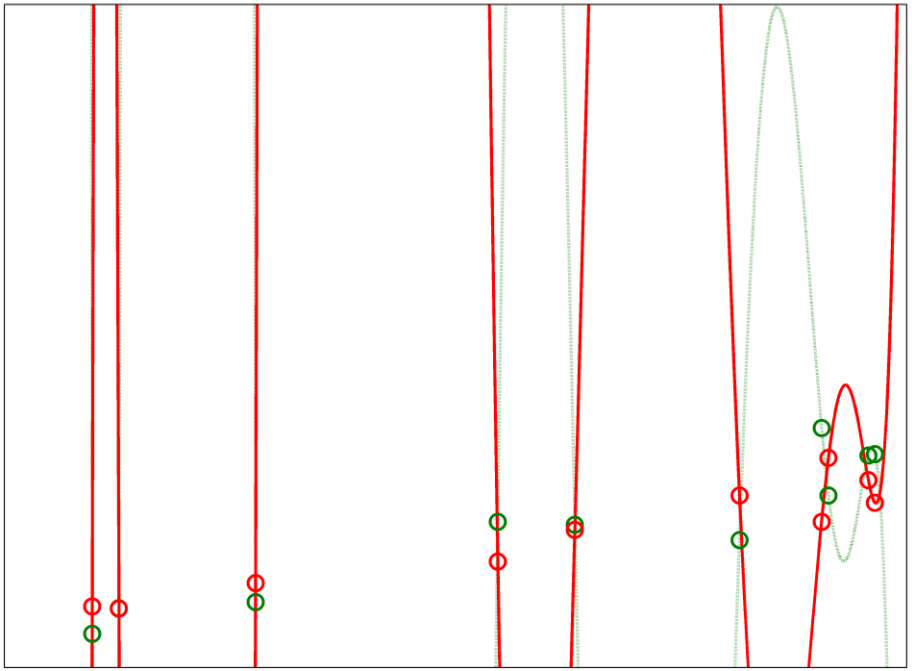


# Simptom peprilagođavanja: velika varijansa

Velika varijansa: male promene trening skupa rezultuju veoma različitim modelima

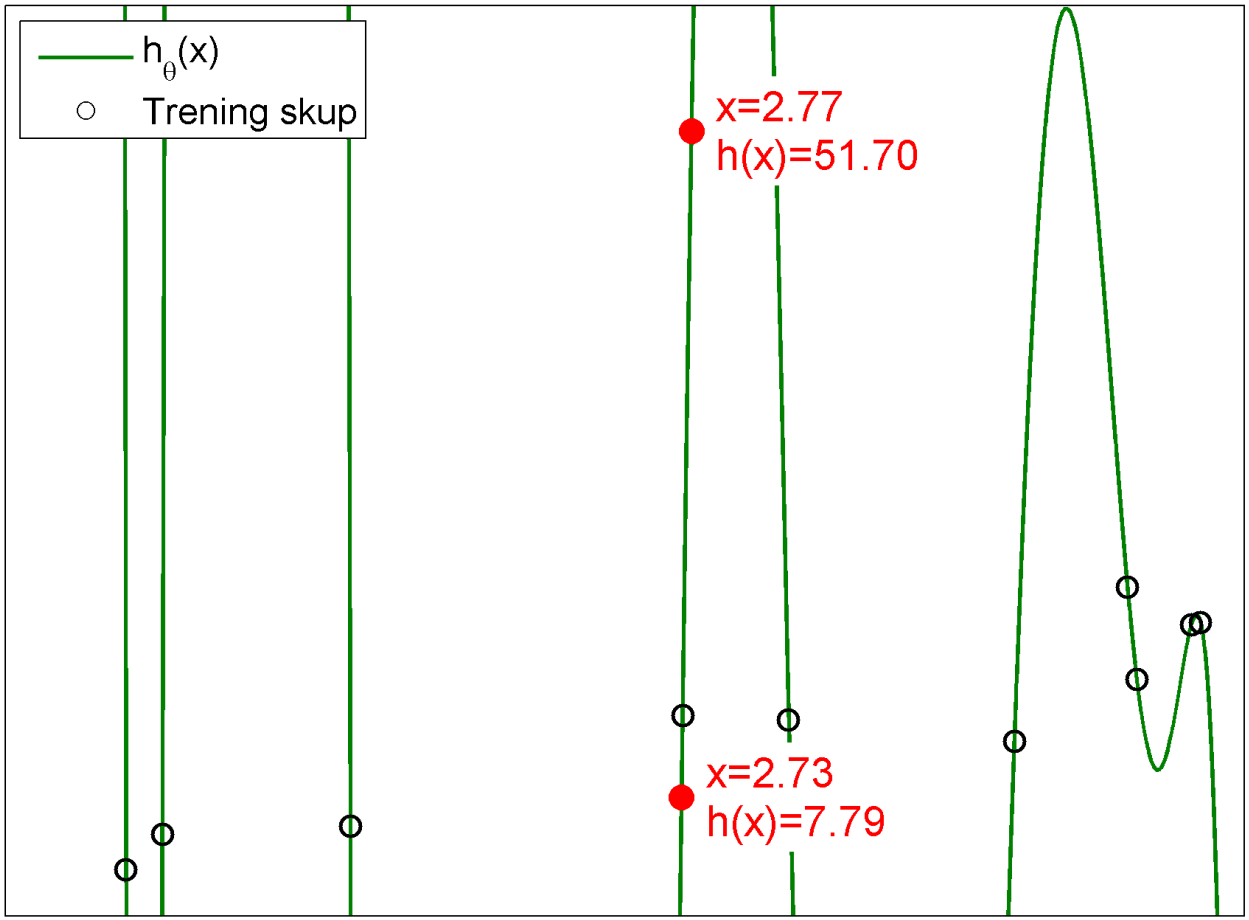


U oba slučaja, $y$ vrednosti su dobijene tako što je dodat šum na stvarnu vrednost $y$.
U oba slučaja je šum jednake jačine

# Simptom peprilagođavanja: velika varijansa

U slučaju velikih vrednosti $\theta$, male promene vrednosti $x$ dovode do velikih promena vrednosti $h_\theta(x)$



# Kako rešiti preprilagođavanje?

Uvećati *N*

Nije uvek moguće

Smanjiti *D*

- Možda ćemo izbaciti obeležja bitna za rešenje problema
- Kompleksnost modela treba da odgovara kompleksnosti problema, a ne *N*

Regularizacija

# Regularizacija

- Sačuvati sva obeležja, ali smanjiti magnitude vrednosti $\theta$
  - *shrinkage methods*

- Ovo će rezultovati „jednostavnijom“ hipotezom koja je manje podložna preprilagođavanju

- Radi dobro u slučaju kada imamo mnogo obeležja takvih da svako doprinosi (u manjoj meri) predikciji $y$

# Regularizacija

- U slučaju linearnih modela

$$h_\theta(x) = \sum_{d=1}^{D} \theta_d x_d$$

važi:

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

Efekat regularizacije:

ograničavamo gradijent funkcije $h_\theta(x)$ (brzinu promene funkcije $h_\theta(x)$ sa promenom $x$)

# Regularizacija

- U opštijem smislu, regularizacijom se naziva bilo koja modikacija optimizacionog problema koja ograničava prilagodljivost modela i čini ga manje podložnim preprilagođavanju

- U još opštijem smislu, regularizacija je bilo kakva modifikacija matematičkog problema koja ga čini bolje uslovljenim