

LASSO i Ridge

- LASSO radi najbolje kada model sadrži mnogo beskorisnih varijabli

$$size = \theta_0 + \theta_1 \cdot weight + \theta_2 \cdot diet + \theta_3 \cdot sign + \theta_4 \cdot speed_{swallow}$$

- Ridge radi najbolje kada je većina varijabli u modelu korisna

$$size = \theta_0 + \theta_1 \cdot weight + \theta_2 \cdot diet + \theta_3 \cdot age + \theta_4 \cdot speed_{swallow}$$

Microarray data analysis

- Među genima koji dele iste biološke putanje se često javljaju velike korelacije
 - o ovakvim genima razmišljamo kao o grupi
- Tipičan cilj je analiza koekspresije gena
 - Geni sa sličnim funkcijama su često u koekspresiji
 - Koekspresija može otkriti dosta toga o regulatornom mehanizmu gena

Cilj

1. Automatska eliminacija trivijalnih gena

 Ridge

 LASSO

2. Automatsko uključenje celu grupu gena u model ako je jedan od gena iz grupe selektovan

 Ridge

Ako su obeležja u velikoj korelaciji, *ridge* će im dodeliti približno jednake koeficijente

 LASSO

Od grupe koreliranih obeležja uključice samo jedno

Rešenje: Elastic Net

- *Elastic Net* je linearna kombinacija *lasso* (L_1) i *ridge* (L_2):

Ridge:
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N L(h_{\theta}(x^{(i)}), y^{(i)}) + \lambda \|\theta\|_2^2$$

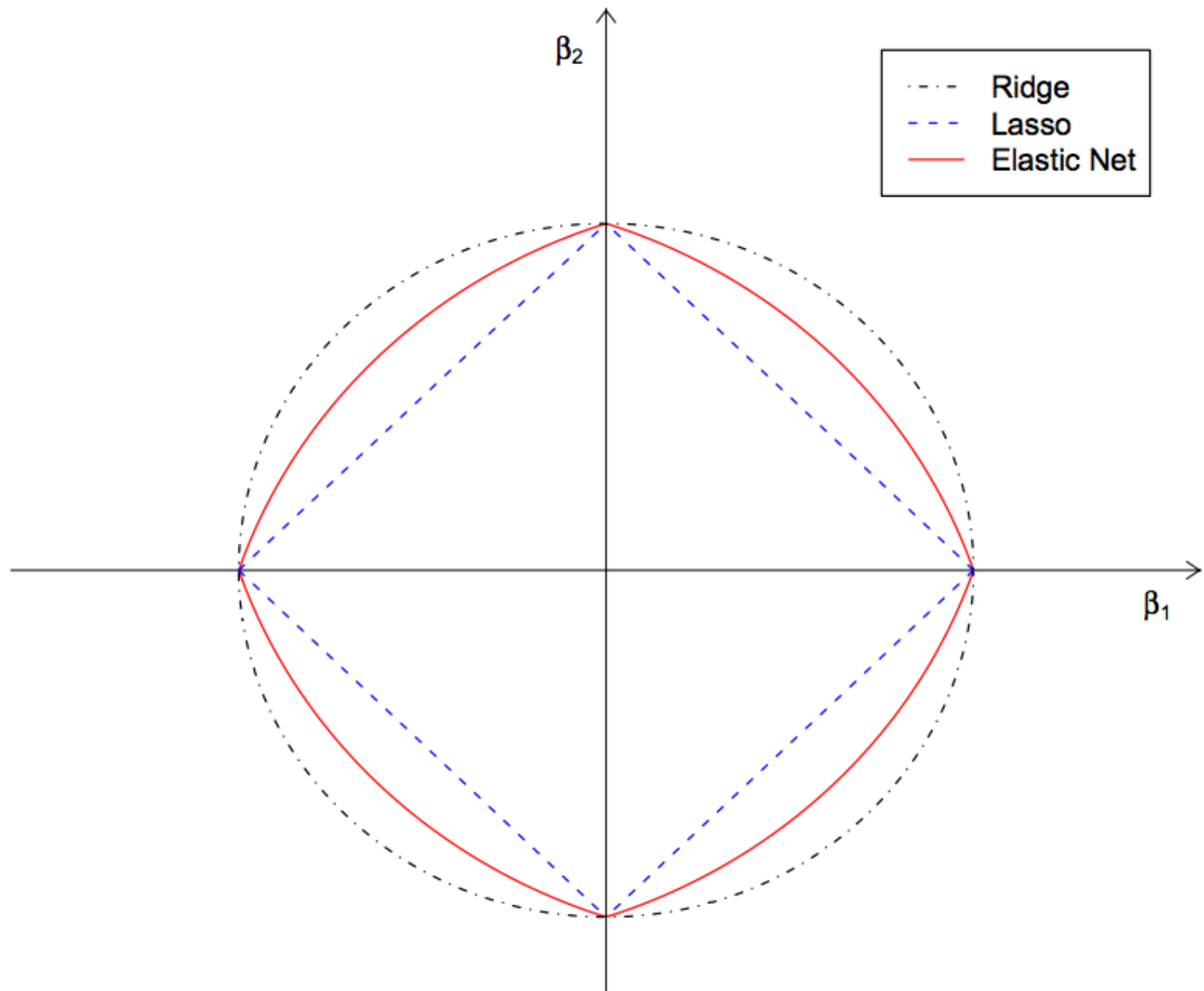
Lasso:
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N L(h_{\theta}(x^{(i)}), y^{(i)}) + \lambda \|\theta\|_1$$

Elastic Net:
$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N L(h_{\theta}(x^{(i)}), y^{(i)}) + \lambda_2 \|\theta\|_2^2 + \lambda_1 \|\theta\|_1$$

- Uklanja ograničenje na broj selektovanih varijabli
 - Ohrabruje grupnu selekciju
 - Stabilizuje L_1 regularizacioni put
- Želimo selekciju obeležja

Elastic Net je predložen u radu: Hui Zou and Trevor Hastie (2005) “Regularization and variable selection via the elastic net”.

Elastic Net



Elastic Net

- *Ridge*, LASSO i *Elastic Net* su varijacije istog *penalty term* oblika:

$$P_{alpha} = \sum_{i=1}^D \left[\frac{1}{2} (1 - \alpha) \theta_i^2 + \alpha |\theta_i| \right]$$

- *Ridge*: $\alpha = 0$
- LASSO: $\alpha = 1$
- *Elastic Net*: $0 < \alpha < 1$

Naive Elastic Net

- Specifikacija $P_{alpha} = \sum_{i=1}^p \left[\frac{1}{2} (1 - \alpha) \theta_j^2 + \alpha |\theta_j| \right]$
se obično smatra za *naive* Elastic Net
- Na žalost, obično ne radi dobro u praksi – parametri se penalizuju dva puta za isti nivo α (stoga se zove naivan)
- Korekcija:

$$P = (1 - \alpha) |\theta|_1 + \alpha |\theta|^2 = \lambda_2 |\theta|^2 + \lambda_1 |\theta|_1$$

- gde je $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$
- $\hat{\theta}_{elastic\ net} = (1 + \lambda_2) \hat{\theta}_{naive\ elastic\ net}$

Sintetički podaci

- Generisaćemo dve nezavisne „skrivenne“ varijable z_1 i z_2 ,
 $z_i \sim U(0,20)$

$$y = z_1 + 0.1 \cdot z_2 + \mathcal{N}(0,1)$$

- Koristićemo sledeće prediktore (šum $\epsilon_i \sim \mathcal{N}(0, 1/16)$):

z_1 grupa

$$x_1 = z_1 + \epsilon_1, \quad x_2 = -z_1 + \epsilon_2, \quad x_3 = z_1 + \epsilon_3$$

$$x_4 = z_2 + \epsilon_4, \quad x_5 = -z_2 + \epsilon_5, \quad x_6 = z_2 + \epsilon_6$$

z_2 grupa

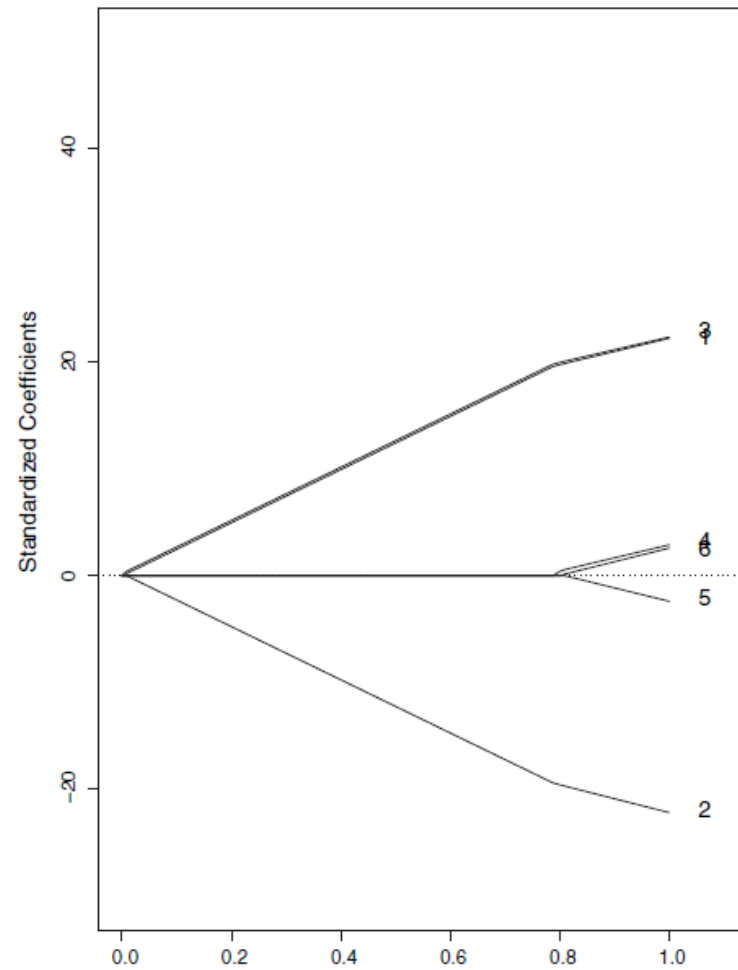
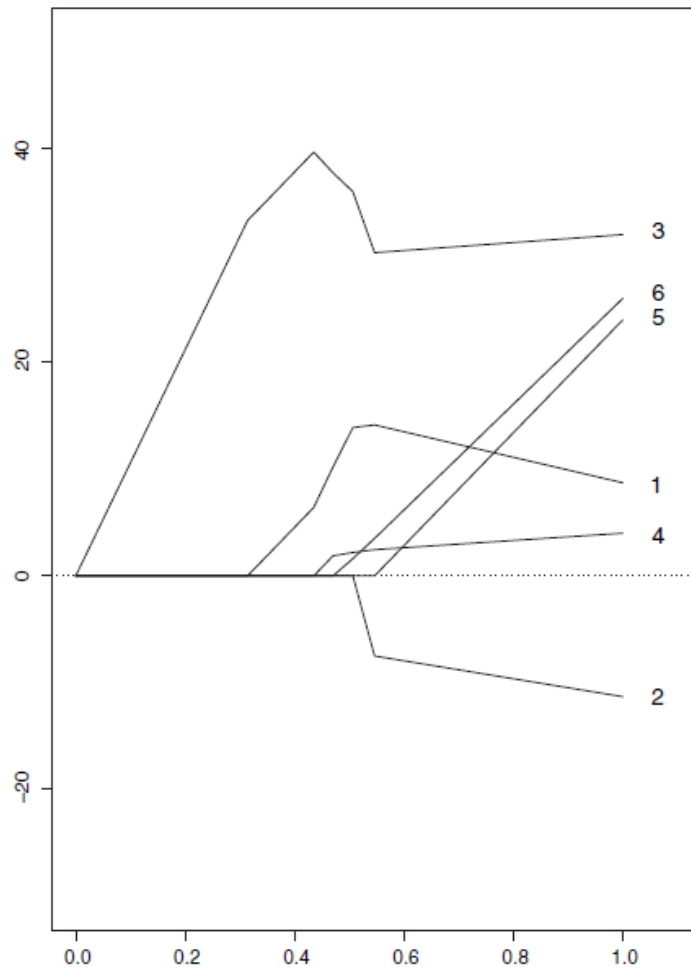
- Na ovaj način je generisano 100 opservacija
- Korelacije unutar grupa su gotovo 1

Sintetički podaci

$$x_1 = z_1 + \epsilon_1, \quad x_2 = -z_1 + \epsilon_2, \quad x_3 = z_1 + \epsilon_3$$

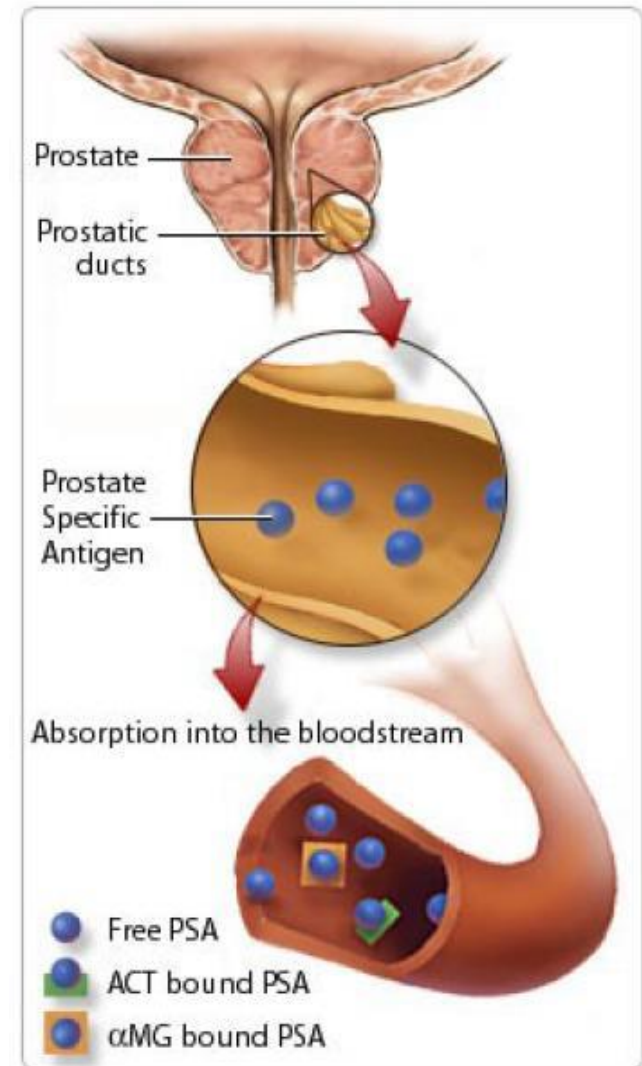
$$x_4 = z_2 + \epsilon_4, \quad x_5 = -z_2 + \epsilon_5, \quad x_6 = z_2 + \epsilon_6$$

$$y = z_1 + 0.1 \cdot z_2 + \mathcal{N}(0,1)$$



Prostate Cancer Data

- Skup podataka dostupan na <https://web.stanford.edu/~hastie/ElemStatLearn/>
- Razmatramo $N = 97$ pacijenata sa rakom prostate i $D = 8$ kliničkih merenja:
 1. log cancer volume (lcanvol),
 2. log prostate weight (lweight),
 3. log benign prostatic hyperplasia amount (lbph)
 4. log capsular penetration (lcp),
 5. age,
 6. Gleason score (gleason),
 7. percentage Gleason scores 4 or 5 (pgg45),
 8. seminal vesicle invasion (svi)
- Cilj: predikcija *prostate-specific antigen* (PSA) – često uvećan kod pacijenata koji imaju rak prostate
- Ciljno obeležje: log of PSA (lpsa)



Prostate Cancer Data

TABLE 3.1. *Correlations of predictors in the prostate cancer data.*

	lcavol	lweight	age	lbph	svi	lcp	gleason
lweight	0.300						
age	0.286	0.317					
lbph	0.063	0.437	0.287				
svi	0.593	0.181	0.129	−0.139			
lcp	0.692	0.157	0.173	−0.089	0.671		
gleason	0.426	0.024	0.366	0.033	0.307	0.476	
pgg45	0.483	0.074	0.276	−0.030	0.481	0.663	0.757

- Vidimo da postoje prilično snažne korelacije među prediktorima
- Na primer, lcavol i lcp
- Za ove varijable očekujemo da su dobri prediktori jer su takođe u snažnoj korelaciji sa ciljnim obeležjem lpsa

Eksperiment

- Poređene su performanse: OLS, ridge, lasso, naive elastic net i elastic net
- Skup podataka je podeljen na:
 - Trening skup koji se sastoji od $N = 67$ opservacija
 - Test skup koji se sastoji od $N_{test} = 30$ opservacija
- Određivanje koeficijenata θ i određivanje parametara modela su izvršeni pomoću desetostruke unakrsne validacije nad trening skupom
- Mera performansi: MSE (*Mean Squared Measure*)

Eksperiment i rezultati su preuzeti iz rada: Zou, H. and Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), pp.301-320.

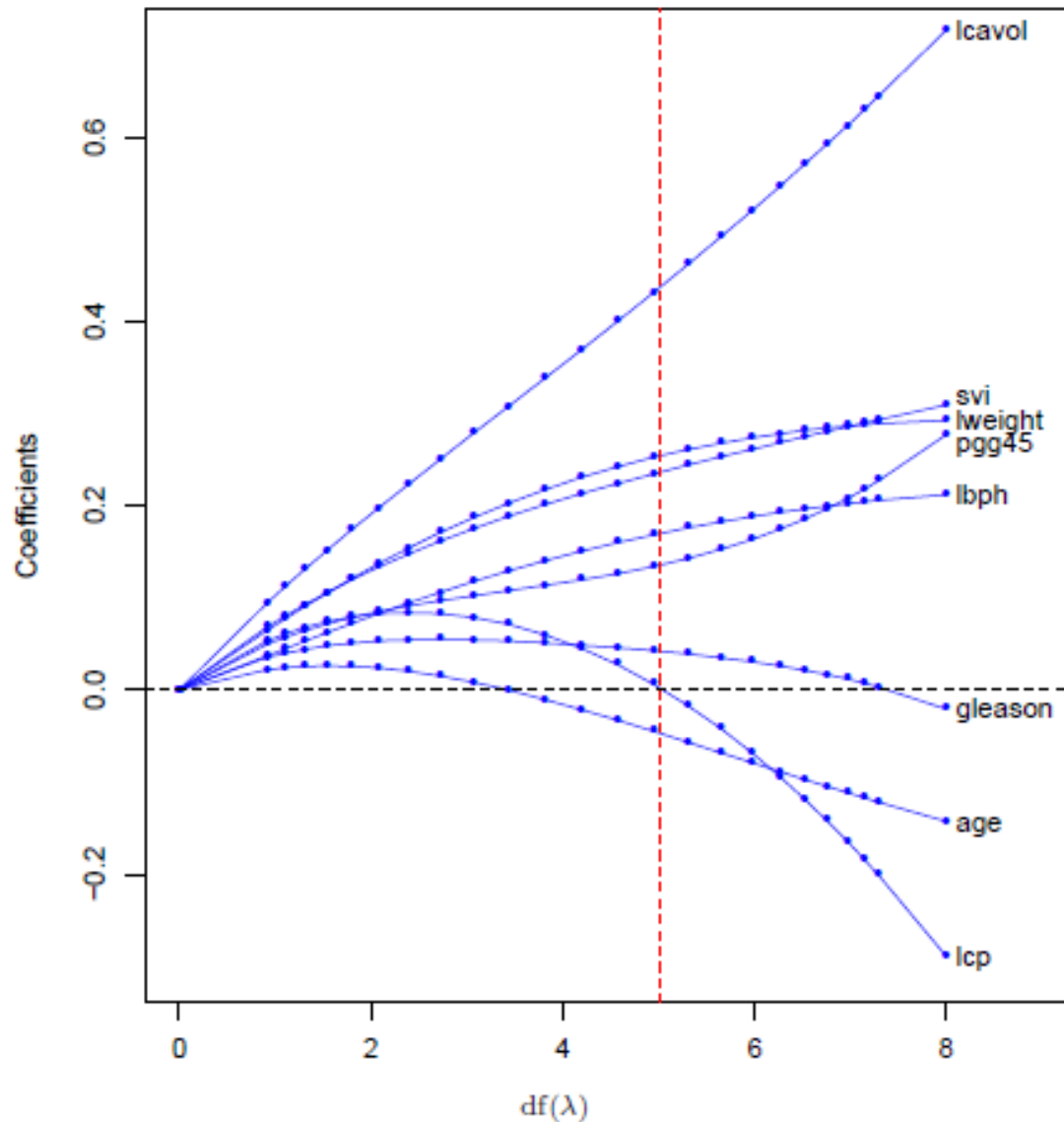
Rezultati

Table 1. Prostate cancer data: comparing different methods

<i>Method</i>	<i>Parameter(s)</i>	<i>Test mean-squared error</i>	<i>Variables selected</i>
OLS		0.586 (0.184)	All
Ridge regression	$\lambda = 1$	0.566 (0.188)	All
Lasso	$s = 0.39$	0.499 (0.161)	(1,2,4,5,8)
Naïve elastic net	$\lambda = 1, s = 1$	0.566 (0.188)	All
Elastic net	$\lambda = 1000, s = 0.26$	0.381 (0.105)	(1,2,5,6,8)

- Najbolje performanse od svih modela je ostvario elastic net
- Najgore performanse je imao OLS metod
- Naive elastic net i ridge su izjednačeni po performansama
- Autori smatraju da su lošije performanse laso posledica velikih korelacija među prediktorima

Coefficient path - Ridge



Poređenje lasso i elastic net

