

# Predviđanje pojave boli u lumbarnom dijelu kralježnice strojnim učenjem

Grupa Sarsorama - Iva Tutiš, Fran Borić, Dario Bogović

## Abstract

Ovaj rad je nastao u nadi boljeg predviđanja, a kasnije i prevencije, boli u lumbarnom dijelu kralježnice. Cilj je sa što većom točnošću, na temelju svojstava poput položaja i kuteva kralježaka, zdjelice, ali i krsne kosti, predvidjeti mogućnost pojave boli, a time i omogućiti planiranje uputstava pacijentu kako spriječiti ili smanjiti bol u lumbarnom dijelu kralježnice.

## 1 Uvod

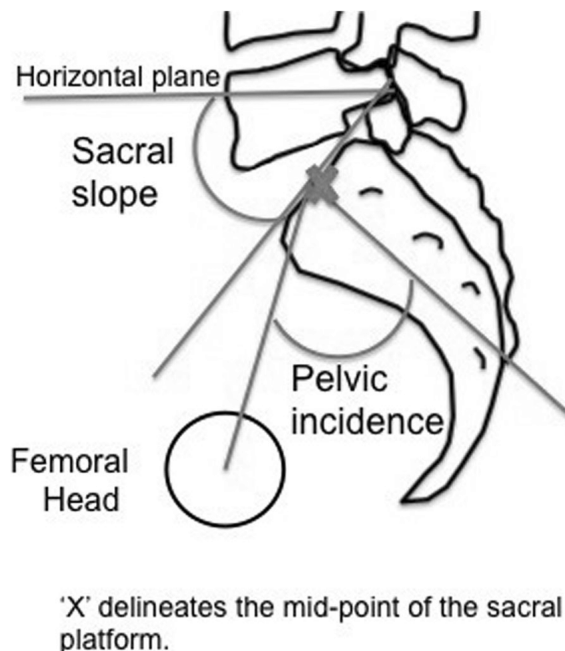
U današnje doba susrećemo se sa sve češćim problemom boli u lumbarnom dijelu kralježnice. Uzroci za to su mnogi, jedan od modernijih je sjedilački način života i prekomjerna težina, ali tu su naravno i starost te različite bolesti i deformiteti (artritis, skolioza). Napretkom tehnologije su se počeli shvaćati fizički razlozi i promijene u kralježnici koje uzrokuju bolove, a sada uz nove pristupe strojnog učenja smo u mogućnosti predvidjeti, uz određenu točnost, bol i prije nego se pojavi. Naša implementacija nam stoga daje jednostavan da-ne odgovor na to pitanje dok klasifikaciju problema prepuštamo drugim istraživanjima.

## 2 Atributi

Prije svake dobre implementacije prvo dolazi određivanje i analiza atributa. Naš dataset se sastoji od njih 12 koje ćemo redom pobrojati i objasniti.

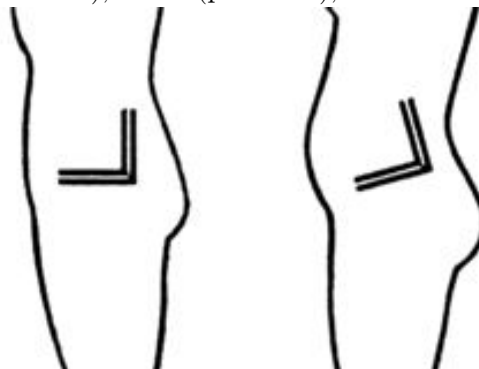
### 2.1 Incidencija zdjelice (Pelvic incidence)

Incidencija zdjelice se definira kao kut između pravca okomitog na središte krsne kosti i pravca koji spaja tu točku s osi bedrene glave bedrene kosti. Detaljnije na slici.



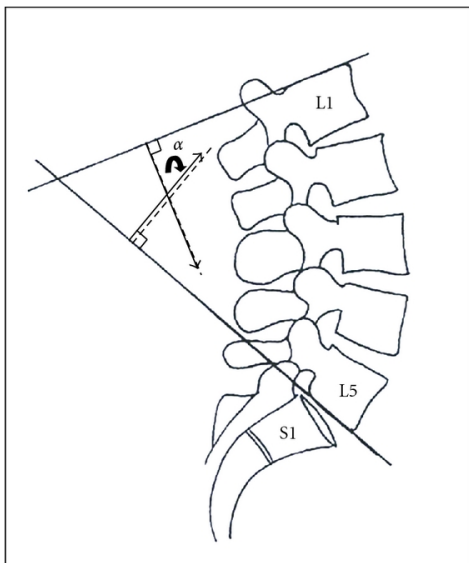
### 2.2 Nagnutost zdjelice (Pelvic tilt)

Kut nagnutosti zdjelice je orijentacija zdjelice u odnosu na bedrene kosti i ostatak tijela. Nagnutost može biti prema naprijed (anterior), nazad (posterior), ali i na stranu.



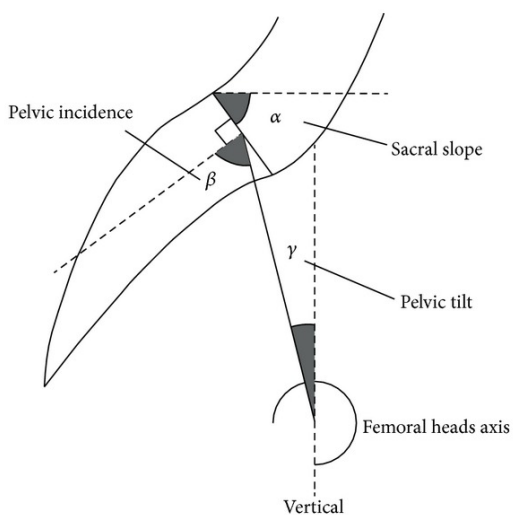
### 2.3 Kut lumbalne lordoze (Lumbar lordosis angle)

Kut za određivanje lordoze, normalna vrijednost je 20-45 stupnjeva.



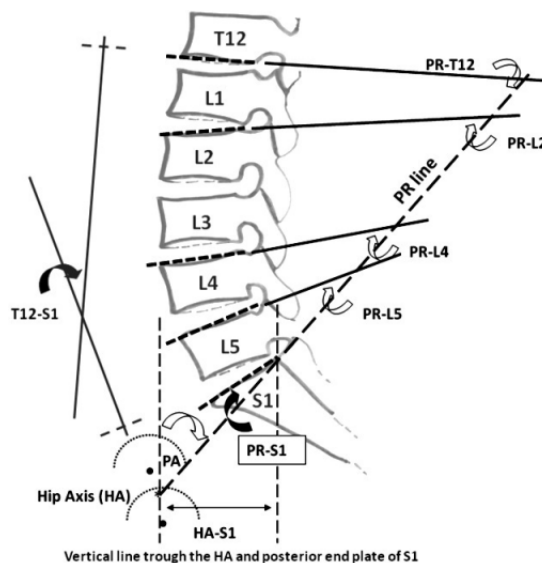
## 2.4 Nagib krsne kosti (Sacral slope)

Nagib krsne kosti (SS) se mjeri između tangente na gornji dio kralješka S1 i horizontalne ravnine.



## 2.5 Zdjelični radijus (Pelvic radius)

Na slici vidimo zdjelični radijus kao PR liniju.



## 2.6 Stupanj spondilolisteze (Degree spondylolisthesis)

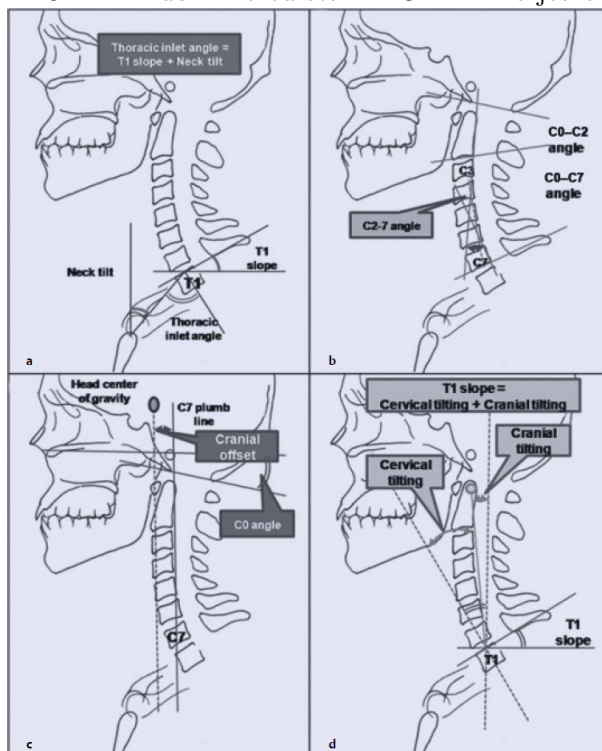
Spondilolisteza je patološki entitet gdje dolazi do subluksacije jednog kralješka u odnosu na drugi, a najčešće se nalazi u području L5/S1 i L4/L5.

## 2.7 Torakalni nagib (Thoracic slope)

Iskrivljenje grudnog dijela kralježnice.

## 2.8 Vratni nagib (Cervical tilt)

Kut između vertikalnog pravca od središta T1UEP i pravca od središta T1UEP do središta C2 kralješka.



## 2.9 Kut krsne kosti (Sacrum/sacral angle)

Sakralni kanal kao nastavak kičmenog kanala prolazi kroz veći dio krsne kosti. Sakralni kut je kut formiran konjugatom s dva dijela krsne kosti. Normalno je veći od 60 stupnjeva.

## 3 Analiza glavnih komponenti - PCA

Kako smo u početku imali 12 atributa nad kojima bi trebali vršiti analizu i testiranja prvo smo napravili normalizaciju podataka i nakon toga nad njima izveli PCA analizu kako bi u najboljem slučaju mogli smanjiti dimenziju na nekoliko stavki i vršiti jednako dobro (ako ne i bolje) treniranje programa nad njima. Naši podaci za PCA analizu sastoje se od 310 redaka normaliziranih podataka, u 12 stupaca. Prvo smo podatke transponirali i nakon toga konstruiramo vektor srednjih vrijednosti (mean vector) koji koristimo u računu matrice raspršenja. Iako

Figure 1: Vektor srednjih vrijednosti

```
[[ -8.02225669e-17]
 [ 2.06286601e-16]
 [ 1.60445134e-16]
 [-1.03143300e-16]
 [ 3.78192101e-16]
 [-2.29207334e-17]
 [-2.86509168e-18]
 [ 8.88178420e-17]
 [ 2.52128068e-16]
 [-4.18303385e-16]
 [-9.16829336e-17]
 [-1.14603667e-16]]
```

je za račun svojstvenih vektora i pripadnih vrijednosti u PCA sasvim svejedno vrši li se račun uz pomoć matrice raspršenja ili matrice kovarijanci, radi kontrole greške smo ozračun napravili na oba načina. Naravno nakon računa uzimamo,

Figure 2: Svojstvene vrijednosti

	Svojstvene vrijednosti matrice raspršenja	Svojstvene vrijednosti matrice kovarijanci
1	1016.557107137095	3.299520825686355
2	1.431034191577671e-13	-2.168615721358713e-16
3	97.04574671339246	0.31406390522133587
4	145.0721446908357	0.46948914142341786
5	386.1177943825857	1.2495721501054582
6	225.2677544831235	0.7290218591686837
7	354.3313741223015	1.1467034761239527
8	327.21913477206536	1.058961601206655
9	317.4192891341556	1.0272469904017986
10	299.974657365121	0.97079171714081588
11	274.1543507058346	0.8872309084331219
12	276.84064648370804	0.895924422773729

po padajućim svojstvenim vrijednostima, sortirano listu parova svojstvenih vrijednosti i pripadajućih svojstvenih vektora. Iz toga u funkcijama treniranja možemo testirati za koliko ćemo zapravo smanjiti dimenziju našeg prostora a da brzina izvršavanja bude u razumnim okvirima

i točnost bude zadovoljavajuća. Također uspoređujemo skupove podataka prije primijene PCA i nakon primijene PCA. Također zbog provjere točnosti uspoređivali smo rezultate s PCA iz paketa sklearn.

Figure 3: Usporedba skupova vrijednosti

SKUPOVI (normaliziranih) PODATAKA PRIJE I NAKON PRIMJENE HARDKODIRANOG PCA

```
ND1 ND2 ND3 ... ND10 ND11 ND12
0 0.147086 0.501369 -0.665177 ... 1.167129 -1.196587 1.712368
1 -1.245864 -0.748769 -1.453001 ... 1.679551 -0.940325 -0.913941
2 0.484370 0.467932 -0.099262 ... 1.635969 -1.227178 -0.615686
3 0.511390 0.711562 -0.411339 ... -0.176157 -1.345020 -0.652989
4 -0.626648 -0.789693 -1.274745 ... -1.059666 -0.190502 -0.069858

[5 rows x 12 columns]
PC1 PC2 PC3 ... PC10 PC11 PC12
0 -0.174284 -1.208808 0.984397 ... 0.103074 -0.314809 -3.426243e-11
1 -2.210186 0.075289 0.815071 ... -0.603445 -0.278021 1.920178e-11
2 0.170257 -0.181279 1.289299 ... 0.616817 -0.519912 2.340197e-11
3 0.316767 -0.852278 -0.464238 ... 0.010709 -0.389510 -3.987569e-10
4 -1.573705 -0.770620 0.512878 ... -0.490473 -0.082999 5.600299e-11

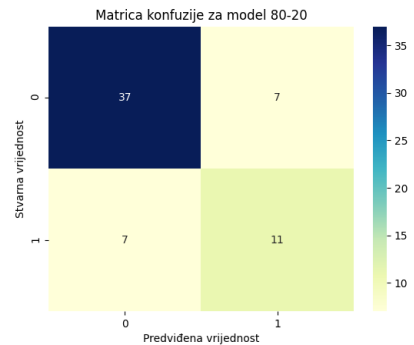
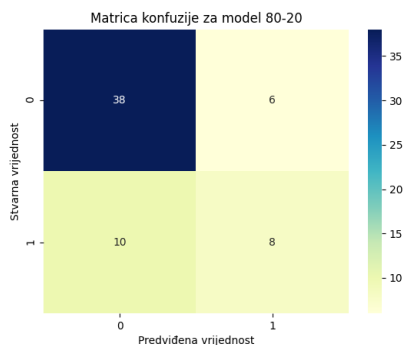
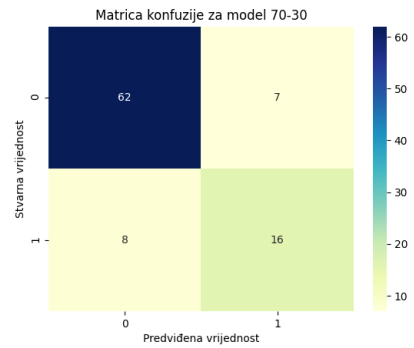
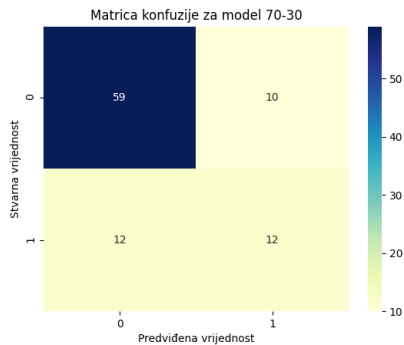
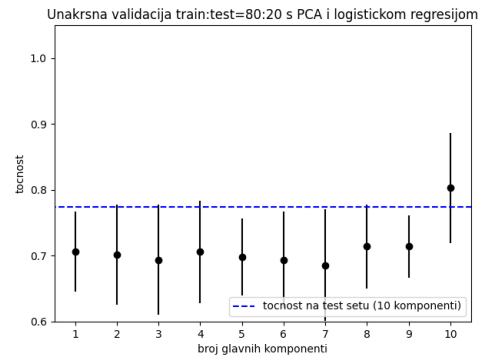
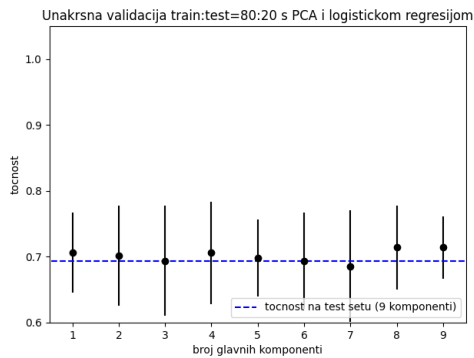
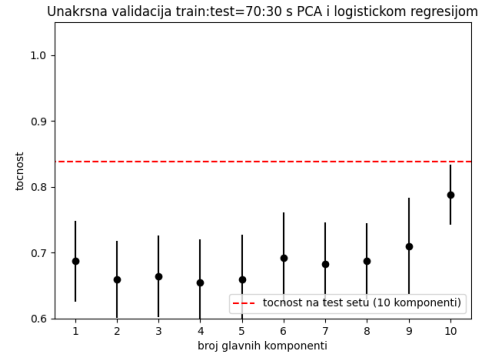
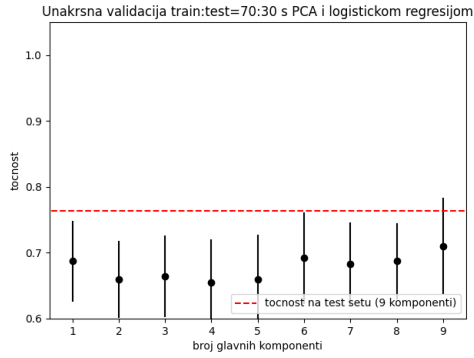
[5 rows x 12 columns]
```

## 4 Pripremanje funkcije treniranja

Za pripremu funkcije treniranja koristit ćemo dva pristupa kako bi mogli testirati koji je bolji. Pristupi se sastoje od toga da nasumično odaberemo 70% podataka u prvom slučaju i 80% podataka u drugom slučaju te na njima treniramo funkciju, dok na ostatku od 30%, odnosno 20%, vršimo testiranje funkcije radi provjere točnosti. Za dane omjere radimo krosvalidaciju, tj. train i test nisu samo jedan skup. Također testiramo s kolikom dimenzijom dobivamo koju točnost. Radimo testiranja na skupovima od 9, 10 i 11 komponenti, sve niže od toga ne da je ni približno zadovoljavajuću točnost. Za klasifikaciju smo koristili algoritam logističke regresije.

### 4.1 Set od 9 komponenti

Na setu od 9 komponenti vidimo znatan pad točnosti, što je i očekivano, ali nedovoljno velik da bi ga se potpuno odbacilo iz analize. Također primjećujemo znatnu razliku u točnosti između modela koji koristi omjer 70-30 (76.34%) i između modela koji koristi omjer 80-20(69.35%) što pripisujemo činjenici da smo prenaučili model koji sam po sebi nema dovoljno komponenti za dovoljnu razinu točnosti.



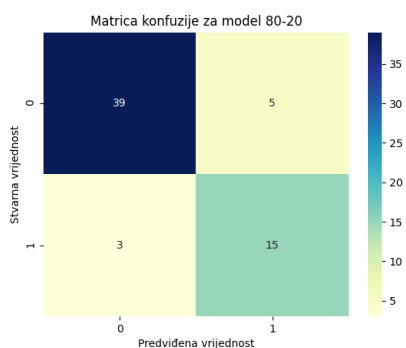
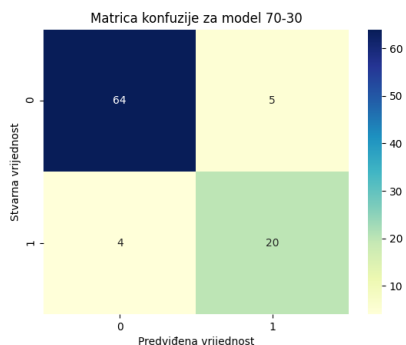
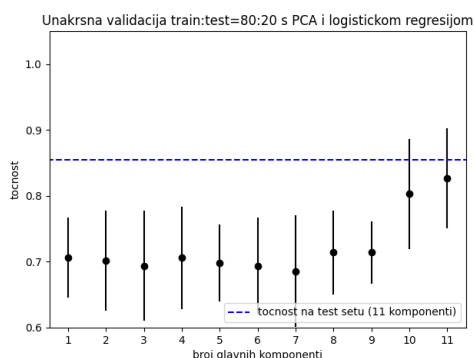
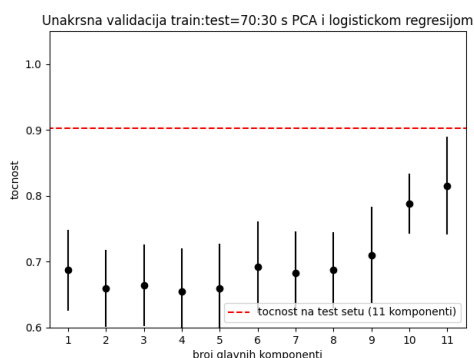
## 4.2 Set od 10 komponenti

Na setu od 10 komponenti vidimo rast u točnosti, ali ponovno izrazito veliku razliku točnosti između modela koji koristi omjer 70-30(83.87%) i između modela koji koristi omjer 80-20(77.42%).

## 4.3 Set od 11 komponenti

Konačno odlučili smo se za set od 11 komponenti. Na njemu primijetimo najveću razinu točnosti od 90.32% za model koji koristi omjer 70-30 i 85.48% za model koji koristi omjer 80-20 što je također i najmanja razlika između ta dva omjera u ovisnosti o broju komponenti. Također kao zanimljivost primijetimo sličnu

točnost između modela koji koristi omjer 80-20 za  $n$  komponenti i između modela koji koristi omjer 70-30 za  $n - 1$  komponenti.



## 5 Zaključak

Dosada mnogi radovi zaobilaze pristup koji smo primijenili što se može vidjeti ovdje. Zbog toga što se često izbjegavao pristup putem PCA dobili smo veću točnost u binarnoj klasifikaciji,

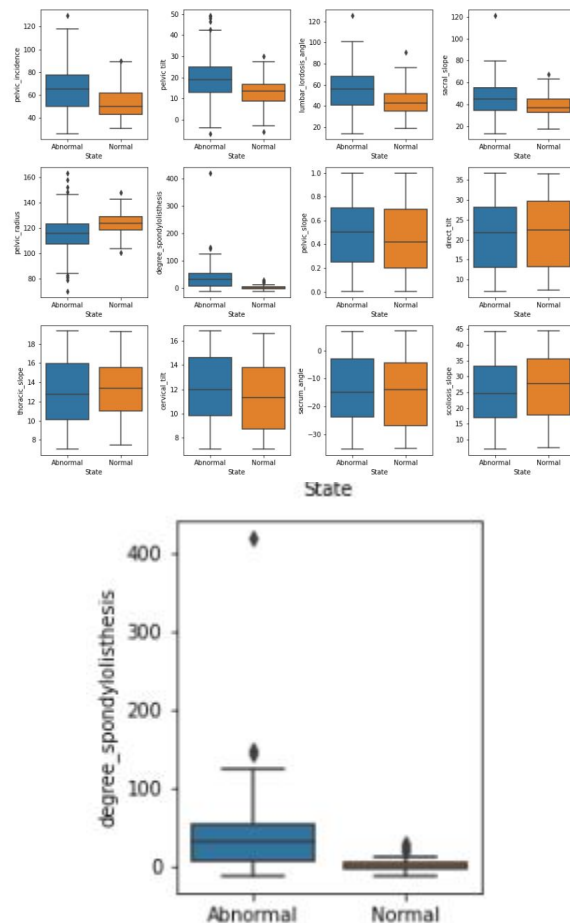


Figure 4: Caption

tj hoće li se razviti bol u leđima ili ne na temelju danog dataseta. Također vidimo da, kako bi održali stupanj točnosti koji želimo vidjeti, možemo spustiti broj komponenti na 11, pa čak i na 10, ali da bi se za veću točnost trebali držati omjera 70-30. Iz eksplorantne analize možemo vidjeti kako je jedna od najbitnijih značajki u razlikovanju abnormalne od normalne kralježnice zapravo stupanj spondilolisteze što vidimo na slici što je dobar smjer za nastavak istraživanja koji bi naveli kao klasifikaciju abnormaliteta abnormalnih kralježnica uz možda detaljniji dataset i medicinske detalje različitih abnormaliteta. Također u slučaju tog smijera istraživanja može se analizirati kako nastalu ili nenastalu bol spriječiti ili smanjiti.

## 6 Reference

[https://github.com/pmf-strojnocucenje/Vjezbe/blob/master/Notebooks/PMF\\_SU\\_2020\\_Vjezbe4\\_00\\_Smanjivanje\\_dimenzionalnosti.ipynb](https://github.com/pmf-strojnocucenje/Vjezbe/blob/master/Notebooks/PMF_SU_2020_Vjezbe4_00_Smanjivanje_dimenzionalnosti.ipynb)  
<https://www.kaggle.com/sammy123/>

lower-back-pain-symptoms-dataset/  
kernels?sortBy=hotness&group=everyone&  
pageSize=20&datasetId=107&outputType=  
all  
[https://web.math.pmf.unizg.hr/nastava/  
su/materijali/](https://web.math.pmf.unizg.hr/nastava/su/materijali/)  
[https://sebastianraschka.com/  
Articles/2014\\_pca\\_step\\_by\\_step.html#  
3-a-computing-the-scatter-matrix](https://sebastianraschka.com/Articles/2014_pca_step_by_step.html#3-a-computing-the-scatter-matrix)  
[https://www.kaggle.com/ashishpatel26/  
lumbago-the-lower-back-pain](https://www.kaggle.com/ashishpatel26/lumbago-the-lower-back-pain)  
[https://en.wikipedia.org/wiki/Low\\_back\\_  
pain](https://en.wikipedia.org/wiki/Low_back_pain)