



**LATVIJAS
UNIVERSITĀTE**

Atbalsta vektoru mašīnas pienesums GRASS GIS attēlu klasifikācijai

Māris Nartišs

GRASS GIS

GRASS ir vecākā (1982.g.) joprojām aktīvi izstrādē esošā brīvā ĢIS programma

Lai arī atbalsta gan rastrus, gan vektorus, laika rindas, 3d rastrus u.tml., tomēr atšķiras no citām populārām ĢIS programmām (QGIS, ArcGIS)

Daļa atšķirību ir vēstures mantojums

Licence: GPL 2.5

Mājaslapa: <https://grass.osgeo.org/>



Klasifikācijas rīki

Pirmā automātiskā tālzpētes attēla klasificēšana ir veikta 1966. g.

GRASS GIS satur automātiskās klasificēšanas rīkus kopš versijas 3.0 1988.g. (iespējams, ka jau versijā 1.1 1986.g.) – maksimālās paticamības klasifikators (*maximum likelihood*; **ML**)

No 1994. gada pieejams *sequential maximum a posteriori* (**SMAP**)

ML un SMAP ir divi vienīgie klasifikatori, kas šobrīd ir iekļauti noklusējuma rīku sarakstā

Abi rīki ir bez būtiskām izmaiņām kopš to radīšanas

No 1996. līdz 1999. gadam bija pieejams arī mākslīgo neironu tīkla klasifikators



Uzlabojumumi versijās 8.0 līdz 8.4

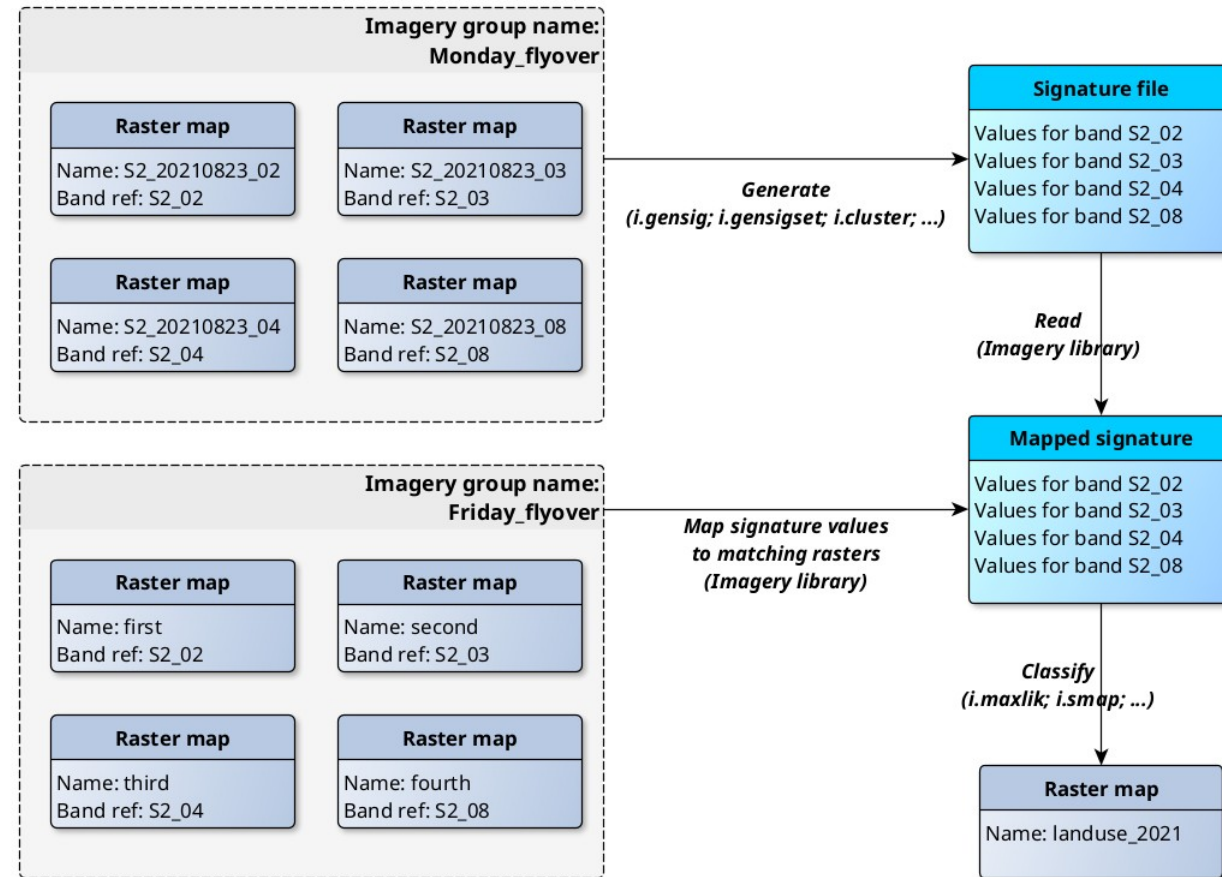
Klasifikatoru spektrālo parakstu faili ir droši pārizmantojami

Jābūt rastra semantisko iezīmju atbilstībai

Kvalitātes kontroles rīka izvade ir mašīnlasāma, kā arī papildināta ar Metjū korelācijas koeficienta vērtību (MCC)

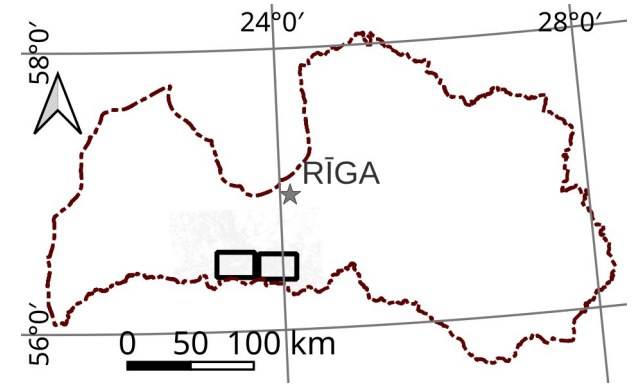
Atbalsta vektoru mašīnas (*support vector machine*, **SVM**) klasifikators

Balstīts uz LIBSVM



Materiāli un metodes

Divas teritorijas Zemgalē ar LAD deklarēto platību poligoniem – viena apmācībai, otra – pārbaudei



ESA Sentinel 2B satelīta 2019. g. 12. jūnija uzņēmums ar četriem kanāliem 10m izšķirtspējā

Pymoo Python bibliotēka optimizēšanai ar daļiņu spieta metodi (*particle swarm optimizer*, **PSO**)

Optimizē SMAP un SVM (C-SVM ar RBF) apmācības hiperparametrus ar inversu MCC kā mērķa rādītāju (mazāks == labāk)

Snieguma vērtējums pārbaudes poligonā

MCC, kopējā un lietotāja pareizība, vizuāls novērtējums

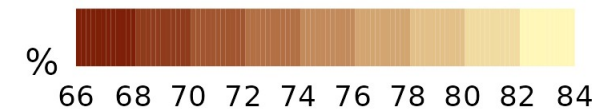
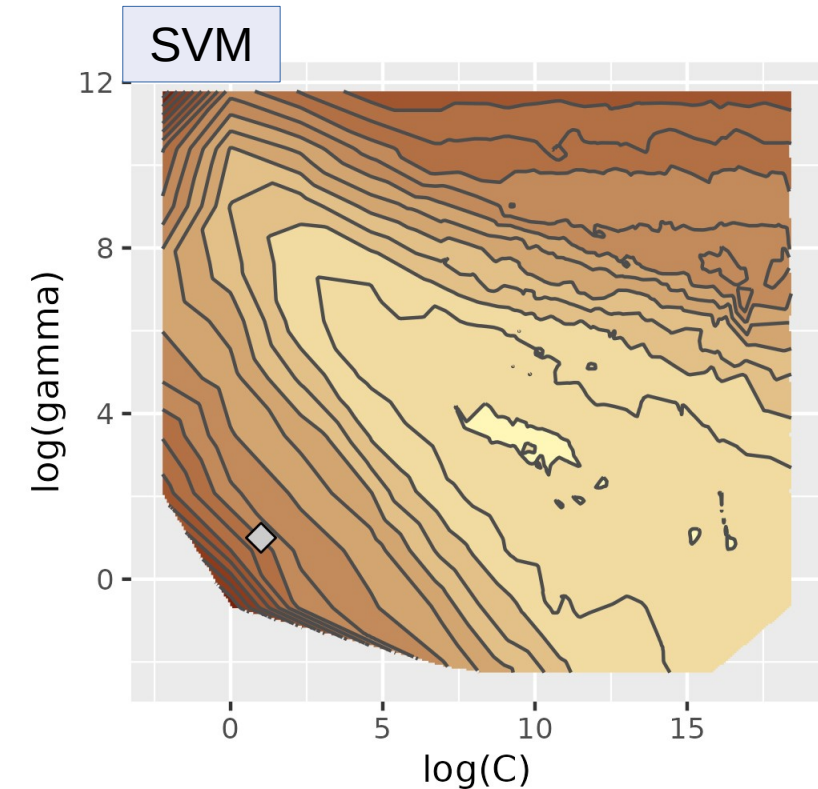
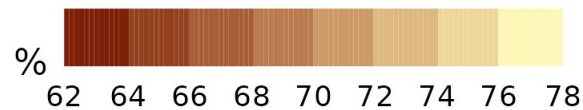
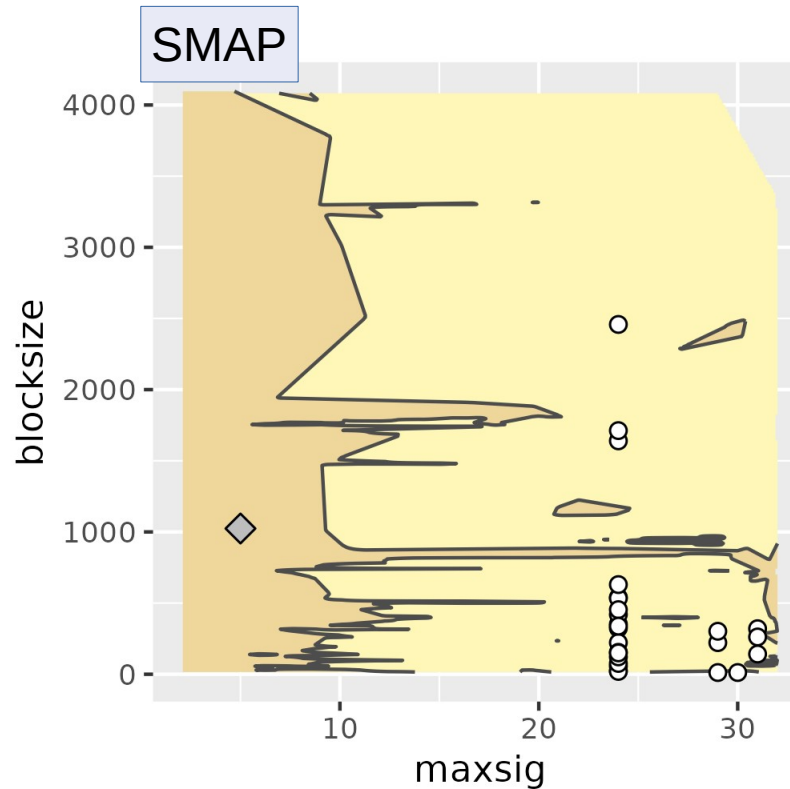
Hiperparametru optimizācija

SMAP nav vērts veikt optimizāciju

Ja ir pietiekams paraugu skaits visās klasēs, ir vērts palielināt katrai klasei noteikto spektrālo parakstu skaitu

Apmācības hiperparametru izvēle būtiski ietekmē SVM rezultātus

Pat manuāla dažu parametru pārbaude var dot būtisku ieguvumu



Pareizības salīdzinājums

Visi klasifikatori uzrāda
pieņemamu sniegumu

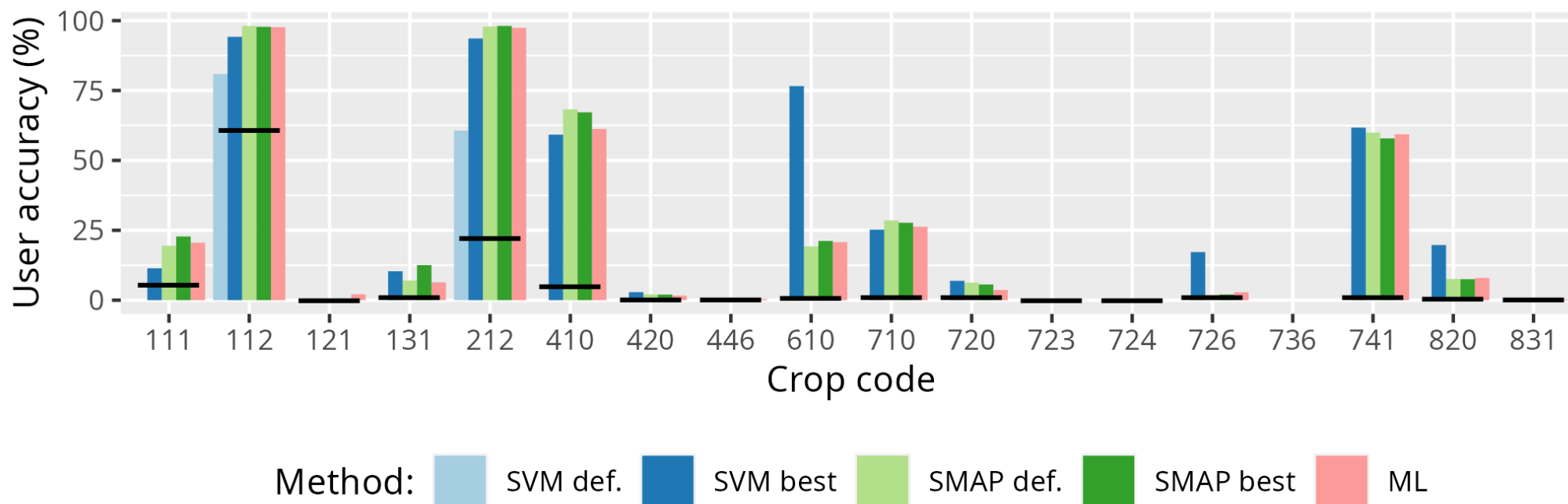
SVM > SMAP > ML

MCC vērtības SVM ar
noklusējuma
hiperparametriem norāda
uz problēmu

Klasifikators	MCC	Kopējā pareizība %
ML	0.622	76.5
SMAP noklus.	0.667	79.6
SMAP labākais	0.675	80.4
SVM noklus.	0.518	73.7
SVM labākais	0.742	85.4

Cik lietojams ir rezultāts?

Divas no izplatītākajām klasēm aptver 82% no testēšanas datu kopas
Neoptimizēts SVM ir ignorējis lielāko daļu klašu (tādēļ MCC zems)!
Nesabalansēts paraugu sadalījums pa klasēm visiem ir „ciets rieksts“

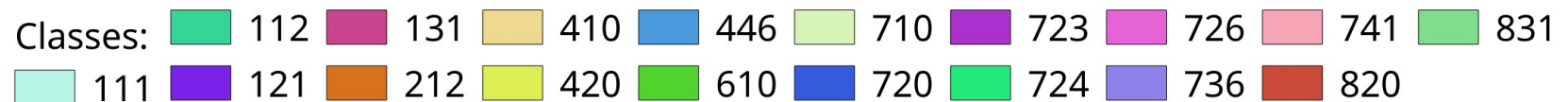
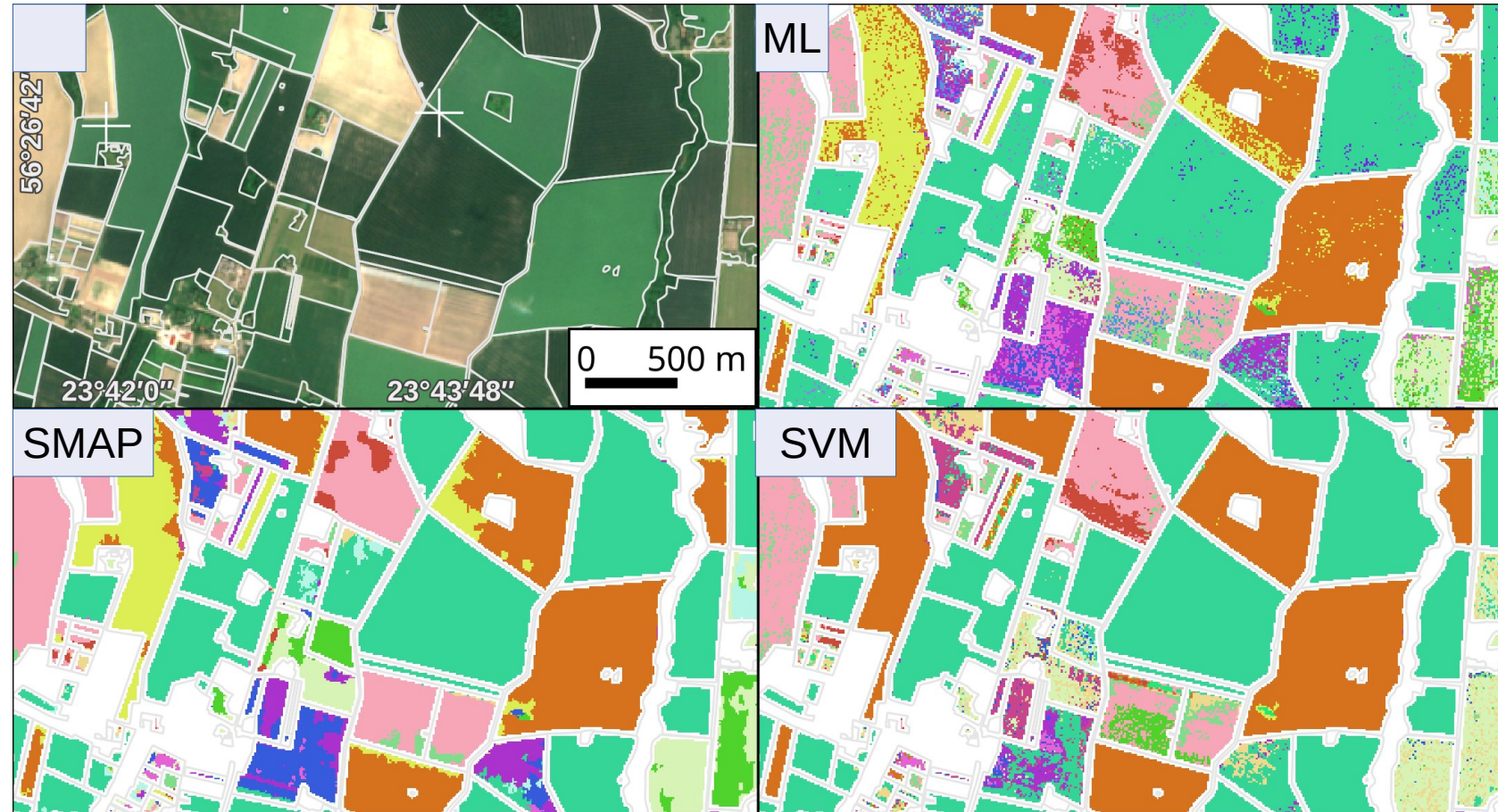


Vizuāls vērtējums

Gan ML, gan SVM
cieš no pikseļu
bāzētajiem
klasifikatoriem
raksturīgā trokšņa
(sāls un pipari)

ML un SVM ieteicams
kombinēt ar
priekš-/pēcapstrādi

SMAP sniedz vizuāli
patīkamāku rezultātu



Veiktspēja

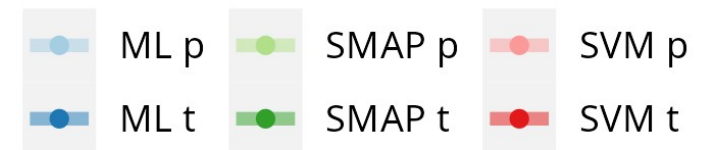
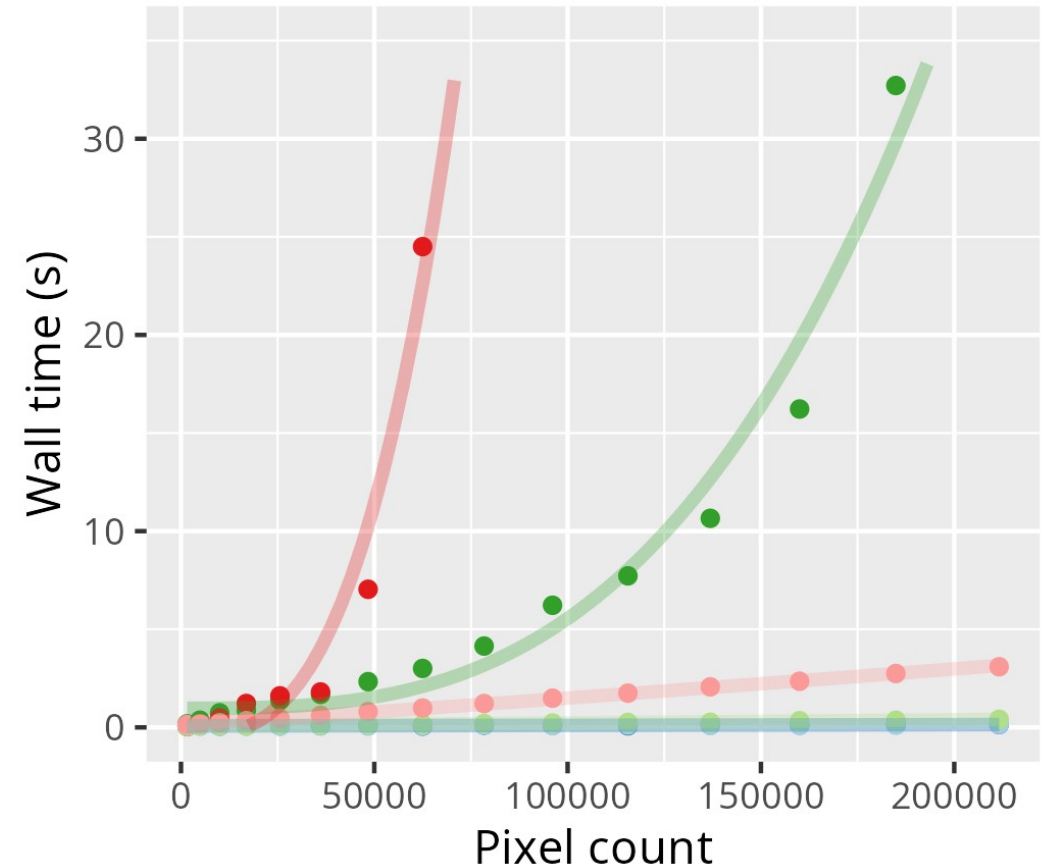
ML klasifikatoram — $O(n)$

SMAP un SVM prognozēšanas fāzei — $O(n \times \log(n))$

SMAP apmācības fāzei — $O(n^3)$

SVM apmācības fāzei — $O(n^{3,5})$

Gan SMAP, gan SVM apmācība nelielam paraugu skaitam ir ar $O(n \times \log(n))$ veiktspēju – vieta uzlabojumiem



Secinājumi

SVM sniedz labus rezultātus, taču datu kopai jābūt labi balansētai

Hiperparametru optimizācija priekš SVM ir obligāta

Neviens no pārbaudītajiem pikseļu bāzētajiem klasifikatoriem nespēj klasificēt lielāko daļu no kultūraugiem no vien Sentinel 2 satelītuzņēmuma

Tikai divām kultūrām lietotāja pareizība pārsniedza 90%

Jāizmanto priekš-/pēcapstrāde, laika rindas, objektu bāzētās klasifikācijas metodes, ...?



Nākotnes perspektīvas

Eksistē tilti starp GRASS un R, scikit-learn, taču tas ir sarežģītāk un lēnāk

No „klasiskiem“ klasifikatoriem vēl iztrūkst gadījuma meža (*random forest*) klasifikators

Dziļo neironu tīklu klasifikators būtu vislabāk, taču tam ir daudz izaicinājumu

Nepieciešami labi priekš-/pēcapstrādes rīki (inteliģenta nogludināšana, labāka poligonu izdalīšana, balsu skaitīšana)

Paldies par uzmanību!

Train a SVM

Obligāti

Do not use the shrinking heuristics (s)

Train a SVC or SVR model for probability estimates (p)

SVM parameters

Degree in kernel function (derīgs diapazons 0-9999): (degree=value)

3 - +

SVM options

Gamma in kernel function: (gamma=value)

1

coef0 in kernel function: (coef0=value)

0

Izvēles

Tolerance of termination criterion: (eps=value)

Komandas izvade

Cost of constraints violation: (cost=value)

1

Rokasgrāmata

The parameter nu of nu-SVC, one-class SVM, and nu-SVR: (nu=value)

0.5

The epsilon in epsilon-insensitive loss function of epsilon-SVM regression: (p=value)

0.1

Close Darbināt Kopēt Palīdzība

Aizvērt dialogu, kad darbs ir pabeigts

Enter parameters for 'i.svm.train'

Pētījums daļēji finansēts no LU zinātnes bāzes (snieguma) finansējuma projekta "Dabas resursu ilgtspējīga izmantošana klimata pārmaiņu kontekstā"



LATVIJAS
UNIVERSITĀTE

111|218|Kvieši, vasaras
112|2355|Kvieši, ziemas
121|2|Rudzi
131|40|Mieži, vasaras
212|884|Rapsis, ziemas
410|188|Lauka pupas
420|4|Zirņi
446|4|Graudaugu un zirņu vai vīķu maisījums ar stiebrzāļu vai
tauriņziežu pasēju, kur proteīnaugi >50%
610|27|Papuve
710|43|Ilggadīgie zālāji
720|46|Aramzemē sēts stiebrzāļu un/vai lopbarības zālaugu (t.sk.
proteīnaugu) maisījums
723|3|Sarkanais āboliņš
726|35|Lucerna
741|31|Citur neminēta kukurūza
820|20|Kartupeļi, kas citur nav minēti
831|3|Lopbarības bietes, cukurbietes

