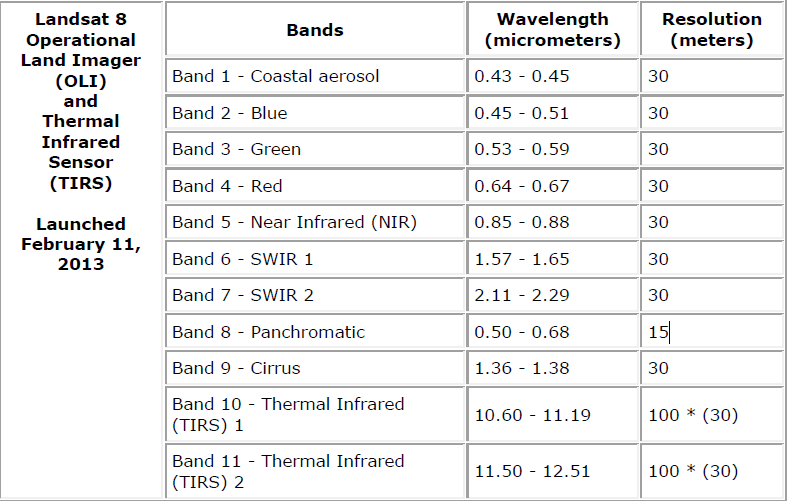
Prepoznavanje oblika

# Tema: Zitarice

# Opis obeležja

Obeležja korišćena u zadatku predstavljaju multispektralna merenja koja su dobijena pomoću satelita LandSat-8. Ovaj satelit za svaki piksel daje 11 multispektralnih merenja od kojih u razmatranje nismo uzimali panhromatski kanal zbog različite rezolucije u odnosu na ostale kanale.

Slika 1

Jedan multispektralni snimak nije bio dovoljan jer različite kulture uspevaju u različito vreme stoga smo morali uzeti četiri multispektralna snimka snimljena u razmacima od mesec dana.

Obeležja koja su korišćena su prvo pretvorena u vrednost reflektanse nakon čega je izvršena dodatna atmosferska korekcija, a dobijene realne vrednosti su zatim reskalirane i kvantizovane na celobrojne vrednosti.

Različita obeležja imaju različit uticaj na tačnost klasifikatora, pa su se u zadatku vršili eksperimenti gde su neka obeležja bila izbačena iz trening i test skupa, neka normalizovana, i izvršeno je poređenje tačnosti klasifikatora za sve navedene slučajeve. U slučaju normalizovanja podataka podaci su normalizovani za svaki sliku posebno, jer slike nisu snimljene u istim uslovima. Normalizacija je izvršena tako što su vrednosti piksela u normalizovanim kanalima svedene na interval [0,1](maksimalna vrednost piksela je 1, a ostale vrednosti se nalaze u zadatom intervalu). U funkciji normalizeData.m može se videti postupak normalizacije.

U zadatku su korišćeni KNN klasifikatori gde je k imalo vrednost {1,2,3,4} i za svaki od gore navedenih slučajeva korišćeni su KNN klasifikatori sa malopre navedenim vrednostima za k. Procenti uspešnosti klasifikacije dati su u sledećim tabelama, a klasifikator koji se pokazao kao najbolji tokom krosvalidacije sa deset podskupova kasnije je korišćen u klasifikaciji neoznačenih piksela test slike. Ceo trening skup je podeljen na deset podskupova. U tabeli 1 dat je raspored piksela po podskupovima.

Tabela . Raspored piksela po klasama u podskupovima

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kukuruz | Pšenica | Soja | Šećerna repa | Suncokret | Ostalo |
| Podskup 1 | 2258 | 1537 | 2950 | 1617 | 1530 | 3112 |
| Podskup 2 | 1422 | 4409 | 830 | 801 | 715 | 4965 |
| Podskup 3 | 3677 | 1661 | 1986 | 1629 | 623 | 3457 |
| Podskup 4 | 5015 | 1474 | 1900 | 1061 | 235 | 3418 |
| Podskup 5 | 2971 | 2228 | 2660 | 605 | 1709 | 2823 |
| Podskup 6 | 2799 | 2926 | 2438 | 1263 | 782 | 2995 |
| Podskup 7 | 2299 | 2420 | 829 | 1591 | 378 | 5537 |
| Podskup 8 | 3549 | 2761 | 1324 | 1247 | 703 | 3480 |
| Podskup 9 | 1747 | 1636 | 1085 | 1893 | 329 | 6319 |
| Podskup 10 | 4173 | 2607 | 981 | 1711 | 3040 | 795 |

Tačnost klasifikatora koji su koristili sva obeležja bez ikakvih transformacija:

Tabela 2. Tačnost klasifikatora u procentima

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | K=1 | K=2 | K=3 | K=4 |  |
| Train1/Test1 u procentima | 76.430 | 75.184 | 75.692 | 75.015 |
| Train2/Test2 u procentima | 75.703 | 75.277 | 75.551 | 74.958 |
| Train3/Test3 u procentima | 71.012 | 71.541 | 71.142 | 71.341 |
| Train4/Test4 u procentima | 76.013 | 74.028 | 75.188 | 74.288 |
| Train5/Test5 u procentima | 88.634 | 88.450 | 89.188 | 88.742 |
| Train6/Test6 u procentima | 78.277 | 77.732 | 79.103 | 79.012 |
| Train7/Test7 u procentima | 69.878 | 69.074 | 70.254 | 69.970 |
| Train8/Test8 u procentima | 69.894 | 70.024 | 70.399 | 70.652 |
| Train9/Test9 u procentima | 82.204 | 81.651 | 82.942 | 82.604 |
| Train10/Test10 u procentima | 69.166 | 69.301 | 69.700 | 69.640 |
| Prosek | 75.721 | 75.226 | 75.916 | 75.622 | 75.916 |

**\*TrainN(skup od devet podskupova celog trening skupa), TestN(jedan podskup trening skupa)**

Tačnost klasifikatora koji su koristili sva obeležja uz normalizaciju poslednja 3 kanala:

Tabela 3. Tačnost klasifikatora u procentima

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | K=1 | K=2 | K=3 | K=4 |  |
| Train1/Test1 u procentima | 80.621 | 80.321 | 81.298 | 80.398 |
| Train2/Test2 u procentima | 79.173 | 79.409 | 78.930 | 79.006 |
| Train3/Test3 u procentima | 80.365 | 79.966 | 79.858 | 79.736 |
| Train4/Test4 u procentima | 82.233 | 81.706 | 81.981 | 81.988 |
| Train5/Test5 u procentima | 88.465 | 88.827 | 89.027 | 88.888 |
| Train6/Test6 u procentima | 81.587 | 81.481 | 82.435 | 82.352 |
| Train7/Test7 u procentima | 77.646 | 75.815 | 77.616 | 77.217 |
| Train8/Test8 u procentima | 76.890 | 77.112 | 77.616 | 77.441 |
| Train9/Test9 u procentima | 81.228 | 80.697 | 81.720 | 81.758 |
| Train10/Test10 u procentima | 74.4495 | 73.630 | 74.118 | 73.728 |
| Prosek | 80.266 | 79.896 | 80.435 | 80.251 | 80.435 |

**\*TrainN(skup od devet podskupova celog trening skupa), TestN(jedan podskup trening skupa)**

Tačnost klasifikatora koji nisu koristili poslednja 3 kanala:

Tabela 4. Tačnost klasifikatora u procentima

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | K=1 | K=2 | K=3 | K=4 |  |
| Train1/Test1 u procentima | 80.713 | 80.375 | 81.274 | 80.413 |
| Train2/Test2 u procentima | 79.188 | 79.462 | 78.968 | 79.044 |
| Train3/Test3 u procentima | 80.380 | 79.966 | 79.866 | 79.728 |
| Train4/Test4 u procentima | 82.423 | 81.904 | 82.126 | 82.118 |
| Train5/Test5 u procentima | 88.488 | 88.850 | 88.981 | 88.881 |
| Train6/Test6 u procentima | 81.579 | 81.473 | 82.405 | 82.329 |
| Train7/Test7 u procentima | 77.646 | 75.869 | 77.639 | 77.217 |
| Train8/Test8 u procentima | 76.890 | 77.120 | 77.349 | 77.457 |
| Train9/Test9 u procentima | 81.251 | 80.805 | 81.781 | 81.812 |
| Train10/Test10 u procentima | 74.487 | 73.690 | 74.201 | 73.750 |
| Prosek | 80.305 | 79.952 | 80.459 | 80.275 | 80.459 |

**\*TrainN(skup od devet podskupova celog trening skupa), TestN(jedan podskup trening skupa)**

U tabelama iznad dati su procenti tacnosti za različite klasifikatore, sada ćemo za specifičan klasifikator(k=3, koristimo Train4-Test4) pronaći sve parametre koje opisuju rad klasifikatora.

Tabela 5. Vrednosti su izražene u procentima.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Kukuruz | Pšenica | Soja | Sećerna repa | Suncokret | Ostalo |
| Osetljivost | 93.273 | 97.649 | 96.25 | 99.749 | 99.252 | 93.241 |
| Stopa lažnih poziva | 4.038 | 5.837 | 4.948 | 0.198 | 1.737 | 3.450 |
| Specifičnost | 95.961 | 94.162 | 95.051 | 99.801 | 98.262 | 96.549 |
| Preciznost | 96.649 | 93.590 | 94.677 | 99.777 | 97.916 | 96.641 |

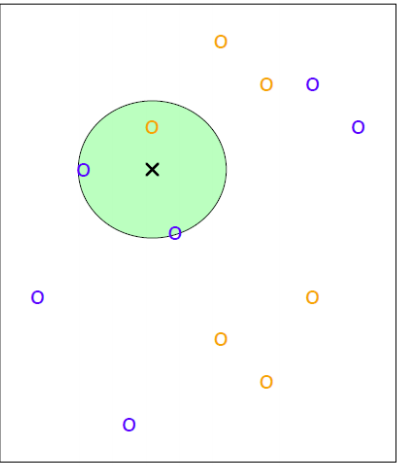
Vidimo da najveći procenat pogodaka imamo kada iz skupa izbacimo poslednja tri kanala. Odatle zaključujemo da su poslednja tri kanala nepotrebna za našu klasifikaciju. Za klasifikaciju neoznačenih piksela koristićemo sva obeležja osim poslednja tri i KNN klasifikator gde je k=3.

## KNN klasifikator

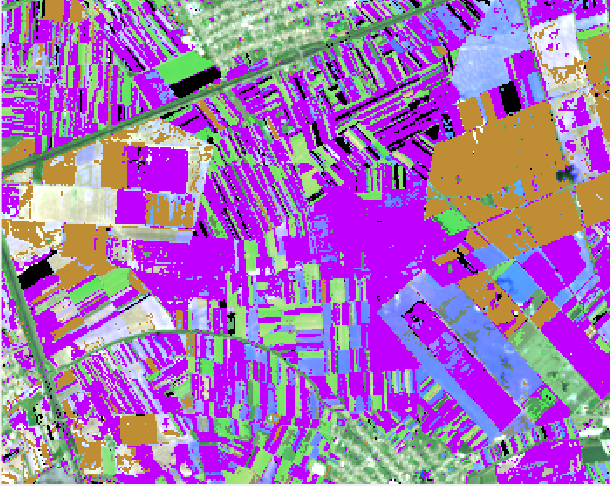
KNN klasifikator spade u kategoriju instance-based metoda mašinskog učenja.To znači da se sve instance iz skupa za trening se čuvaju u memoriji, kada se pojavi nova instanca koju je potrebno klasifikovati pronalazi se k trening instanci koje su najbliže novoj instanci. Klasa nove instance se procenjuje na osnovu k najbižih trening instanci.

KNN posmatra sve instance kao tačke u n-dimenzionalnom prostoru gde je n broj obeležja kojima se instanca opisuje, koristi odgovarajuću metriku za računanje udaljenosti instanci i klasifikuje instancu tako što bira najpopularniju(klasa koja ima najviše članova među susedima instance) klasu medju najbližim susedima te instance. Jedna od mana ovog klasifikatora je što sve instance trening skupa čuva u memoriji. U tom slučaju treba ukoliko radimo sa ogromnim trening skupom, treba ga smanjiti, ali ostaviti isti odnos broja observacija različitih klasa.

Slika 1. Primer KNN kada je k=3



Slika Vizuelni prikaz klasifikacije

****

0) kukuruz-ljubičasta

1) pšenica-zelena

2) soja-plava

3) šećerna repa-braon

6) suncokret-crna