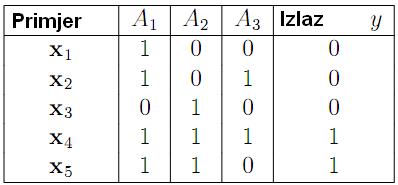
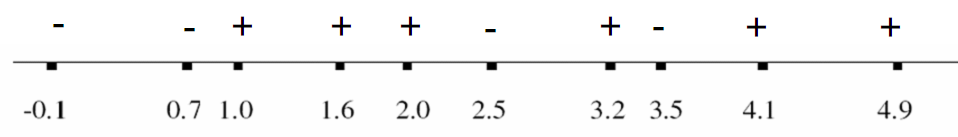
**Mašinsko učenje, četvrti domaći zadatak, decembar 2017.**

1. (Naive Bayes, drvo odlučivanja) Dati su sljedeći podaci, sve tri promjenljive su binarne.
   1. Kreirati drvo odlučivanja po datim primjerima. Prikazati sva izračunavanja. Kako će biti klasifikovani primjeri x6=(0,0,0) i x7=(0,1,1)?
   2. Izračunati sve vjerovatnoće potrebe za Naive Bayes, Kako će biti klasifikovani primjeri x6=(0,0,0) i x7=(0,1,1)?
   3. Ponoviti korak b, koristeći Laplasovo poravnavanje.
2. (Cross validation) Dat je sljedeći skup primjera za obučavanje, sa jednim realnim ulazom x i jednim binarnim izlazom y. Koristićemo k-nn sa Euklidskim rastojanjem.



* 1. Ako koristimo leave-one-out cross-validation, kolika je greška 1-nn klasifikatora na datom skupu? Obrazložiti odgovor.
  2. Ako koristimo leave-one-out cross-validation, kolika je greška 3-nn klasifikatora na datom skupu? Obrazložiti odgovor.
  3. Koji klasifikator birate na osnovu dobijenih grešaka?

1. Primjenom k-means algoritma sa Euklidskim rastojanjem grupisati sljedećih 8 primjera u 3 klastera: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). Početni klasteri su A1, A4 i A7. Uraditi samo jedan prolaz kroz primjere. Prikazati nove centre klastera i na mreži 10x10 nacrtati primjere i označiti klastere. (10)
2. (k-NN, klasifikaciono drvo) Na pjevačkom takmičenju „Studentski idol 2013“ prijavljeni su studenti koji žele da postanu muzički idoli. Dva su kriterijuma za izbor kandidata: radne navike (engl. work ethic ili W) i talenat (engl. raw talent ili R). Klasifikujemo kandidate kao potencijalne superstarove (predstavljeni crnim tačkama) ili kao obične studente (predstavljene kružićima) pomoću nearest-neighbors klasifikatora.
   1. Na slici 1 nacrtati granice koje nalazi 1-NN klasifikator.
   2. Kako će 1-NN klasifikator klasifikovati kandidata na poziciji (5,4)? Obrazložiti.
   3. Kako će 3-NN i 5-NN klasifikatori klasifikovati kandidata na poziciji (5,4)? Obrazložiti.
   4. Sad mijenjamo pristup i prelazimo na klasifikacione drveta. Dva su moguća kriterijuma ta podjelu: R>5 ili W>6. Za oba kriterijuma odrediti prosječnu entropiju. Koji će kriterijum biti odabran?
   5. Pretpostavimo da je algoritam kao prvi kriterijum izabrao R>5. Na slici 2, precizno nacrtati sve ostale granice koje će algoritam postaviti.
3. (Perceptron) Prikazati korake izvršavanja perceptrona za klasifikaciju izlaza binarne funkcije NAND (negacija funkcije AND). Stopa obučavanja je 0.1. Skup za obučavanje je {(<0,0>,1), (<0,1>,1), (<1,0>,1), (<1,1>,1)}. Dodati x0 kao bias. Početni težinski koeficijenti su w0=0.3, w1=0.1, w2=0.1.
4. (Regresija) U tabeli 1 su dati podaci o starosti vozača i daljini u metrima – koliko daleko vozač vidi iz vozila. Odrediti regresionu liniju za dati skup podataka. (Možete napisati program ili koristiti program za tabelarne proračune – Excel ili Calc).
5. (SVM) Na slici 3 dati su članovi dvije različite ekipe u igri Drones. Članovi prve ekipe su označeni štitom, a druge ekipe krugom.
   1. Koji se o sljedećih kernela može upotrebiti za klasifikacju članova ekipa: lineani, kvadratni, RBF ili nijedan od njih? Obrazložiti.
   2. Na slici 4 su prikazani drugi članovi ekipa, koji su sada linearno separabilni. Punom linijom nacrtati granicu, isprekidanom linijom nacrtati tzv. gutter-e i zaokružiti potporne vektore.
   3. Na osnovu granice iz b., odrediti parametre w, b i α. Prikazati sva izračunavanja. Indeksi koeficijenata α treba da budu imena članova ekipe (npr. αA, αH...). Klasifikator treba da vrati +1 za krugove i -1 za štitove.
   4. Na slici 5 su uklonjeni svi takmičari iz elipe krugova, a njihove bivše pozicije su označene sa ***x***. Potrebno je postaviti samo jednog člana ekipe krugova tako da se granica dobijena u b. ne promijeni. Na koje od pozicija 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 ili nijednu od njih je moguće postaviti tog jednog člana? Obrazložiti.

|  |
| --- |
| Age Distance |
| 18 510  20 590  22 560  23 510  23 460  25 490  27 560  28 510  29 460  32 410  37 420  41 460  46 450  49 380  53 460  55 420  63 350  65 420  66 300  67 410  68 300  70 390  71 320  72 370  73 280  74 420  75 460  77 360  79 310  82 360 |

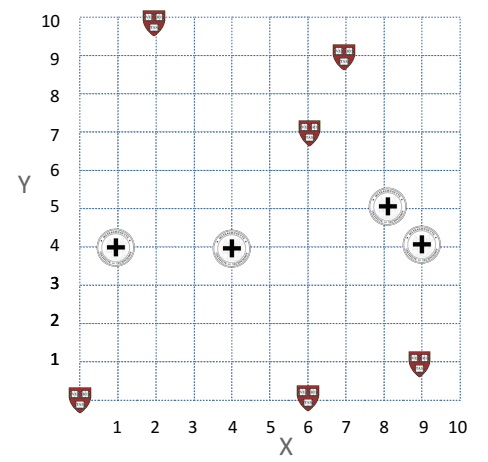
Tabela



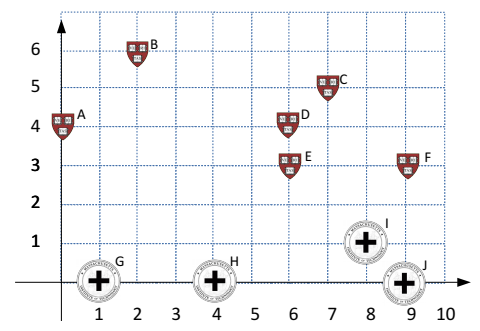
Slika 1



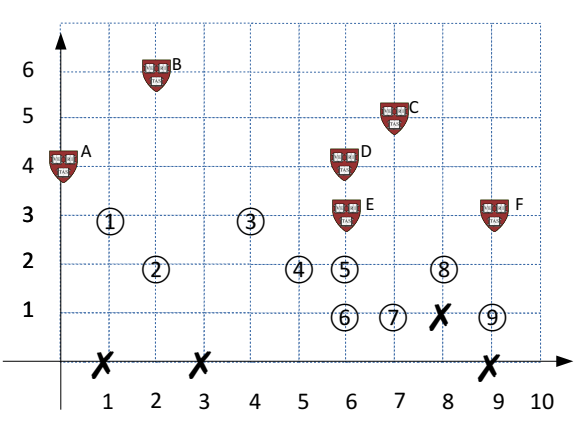
Slika 2



Slika



Slika



Slika

Rok za predaju rada je 09. januar do 12,00. U ovaj fajl, počevši od stranice 6, napisati rješenja zadataka. Ime datoteke mora biti oblika IME\_I\_PREZIME\_DOMACI04.RAR i šalje se kao attachment na imejl ai@rc.pmf.ac.me sa temom (subject-om) poruke "Ime i prezime Domaci 4". Napomena: Radove zipujte primjenom WinRAR-a ili 7Z-a, jer mejl-server PMF može da odbije sve poruke koje u attachment-u imaju datoteke sa ekstenzijom ZIP.