k Test 2 Data Mining 2014

1. Какъв е моделът на следните неща (локален или глобален):  
   индукция на дърво на решения - глобален

naïve bayes - глобален

kNN - локален

невронна мрежа - глобален

1. Кой е коренът на дървото при дадена **(примерна)** таблица:
   1. F1
   2. F2
   3. F3 - output би трябвало да е класът

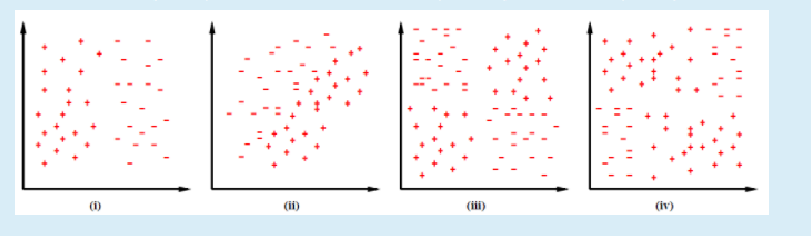
| F1 | F2 | F3 | Output |
| --- | --- | --- | --- |
| a | b | c | + |
| c | a | c | + |
| a | a | c | - |
| c | c | a | - |
| a | b | b | - |

* 1. Output

1. Кой атрибут ще сложите за корен на дървото?
   1. F1
   2. F2
   3. F3
   4. Output

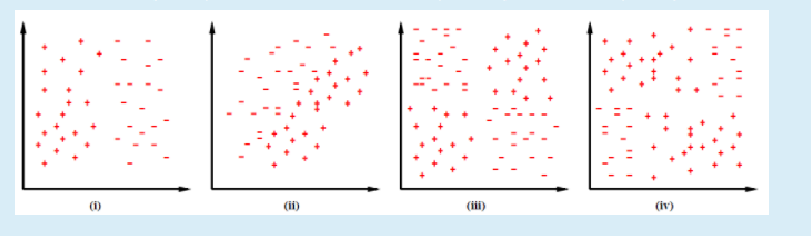
| F1 | F2 | F3 | Output |
| --- | --- | --- | --- |
| a | а | а | + |
| c | b | c | + |
| c | a | c | + |
| a | a | b | - |
| a | b | a | - |
| c | b | b | - |

1. Може ли персептронът да реализира XOR?
   1. да
   2. не
2. Кое твърдение се отнася за ансамбловото учене (Ensemble learning)?  
   a) използват се различни алгоритми и се прави гласуване  
   б) използват се различни извадки от тестови данни и се прави гласуване  
   в) използват се к члена и се гласува
3. Кои от следните алгоритми спадат към йерархичните модели?
   1. kNN
   2. Agglomerative learning
4. Кои от следните алгоритми спадат към ученето с учител (supervised learning)?
   1. ID3
   2. DBSCAN
   3. kNN
   4. kMeans
   5. C4.5
5. Кои от следните твърдения се отнасят за асоциативните правила?
   1. минимален support
   2. минимален confidence
6. Кои от следните не са верни за асоциативното учене?  
   а) резултатът от косинуса на ъгъла между данните
7. Кой от следните алгоритми за обучение се явява мързелив
   1. Индукция на дърво на решенията
   2. учене основано на примери
   3. линейна регресия
   4. k-means
8. Посочете правилното твърдение за обратното разпространение на грешката (Error Backpropagation)
   1. използва се при обучението на еднослойна невронна мрежа (perception)
   2. използва се при обучението на еднослойна и многослойна невронна мрежа
   3. не се използва при обучението на невронни мрежи
   4. използва се при обучението на многослойна невронна мрежа
9. Кое от следните обучаващи се множества не може да бъде разделено на дърво на решенията на 2 нива



* 1. (i)
  2. (ii)
  3. (iii)
  4. (iv)

1. Кое от следните обучаващи се множества може да бъде правилно разделено от персептрона?



Отбележете едно или повече

* 1. (i)
  2. (ii)
  3. (iii)
  4. (iv)

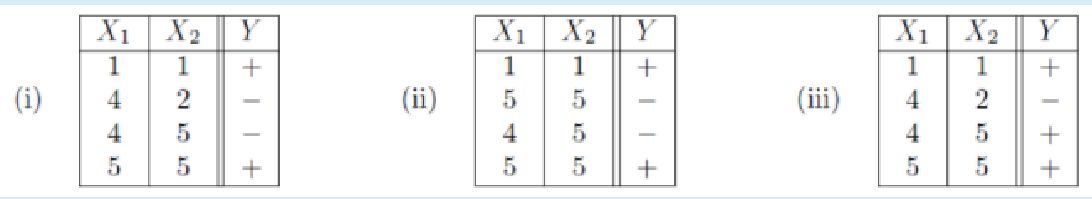
1. При ученето на дърво на решенията, по кой принцип избираме една от множеството допустими хипотези
   1. на Фишер
   2. на Окам
   3. на Тюринг
   4. на Бейс
2. При ученето на дърво на решенията, по кой принцип избираме една от множеството съвместими хипотези
   1. на Марков
   2. на Окам
   3. на Тюринг
   4. на Бейс
3. Каква е основната цел на кастренето (pruning) на дърво на решенията
   1. Да направи дървото по-компактно
   2. Да повишим точността върху обучаващото множество
   3. Да намалим пространството от хипотези, в които търсим
   4. Да предотврати прекомерно нагаждане към обучаващите примери (over-fitting)
4. Изводът чрез изброяване (inference by enumeration) се основава на:
   1. условните автомарности между автомарните събития
   2. пълното съвместно разпределение на автомарните събития
   3. доказателства, получени от автомарните събития
   4. единствено от автомарните събития
5. Ансамблово учене (Ensemble learning) e:
   1. от множество обучаващи се примера се избира представителна извадка (k на брой примера), така наречения парламент, който после се използва за обучение
   2. Учене на няколко класификатора, като се използва един и същи учещ алгоритъм, но с различни тегла на примерите в обучаващото множество. После тези класификатори вземат решение с гласуване.
   3. Учене на няколко класификатора, използвайки различни алгоритми за учене, които вземат решение с гласуване
   4. Намираме k най-близките съседни примери от обучаващото множество, които вземат решение с гласуване
6. Какъв модел учи всеки един от следните алгоритми (глобален (g) или локален (l))
   * 1. Учене основано на примери (kNN)
     2. Наивен Бейсов класификатор
   1. 1 l, 2 l
   2. 1 l, 2 g
   3. 1 g, 2 l
   4. 1 g, 2 g
7. Какъв модел учи всеки един от следните алгоритми (глобален (g) или локален (l))
   * 1. Невронна мрежа
     2. Индукция на дърво на решенията
   1. 1 l, 2
   2. 1 l, 2 g
   3. 1 g, 2 l
   4. 1 g, 2 g
8. Кой от изброените клъстерни алгоритми изгражда йерархичен модел отдолу нагоре
   1. агломеративен
   2. нито един от изброените
   3. йерархичен k-means
   4. k-http://fmi.wikidot.com/ai05means ++
9. Правилото (от началното състояние към целта (Progressive)) и обратно (от целта към началното състояние (Regressive)) търсене се използват само за строене изцяло (totally) наредени планове
   1. истина
   2. лъжа
10. Причинно-следствена връзка (casual link) при частично наредените планове се означава

**A -p-> B**

и означава:

* 1. Между A и B трябва да има друго действие C, което да дава резултат p
  2. A трябва да се изпълни преди B
  3. Изпълнението на B трябва да започне непосредствено след изпълнението на A
  4. Не може да има действие C между A и B, което да отменя p

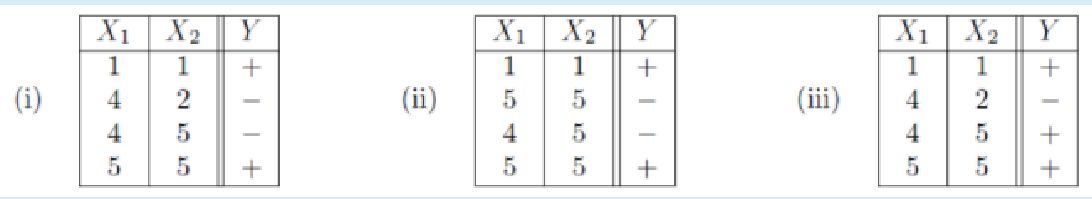
1. Алгоритъмът k-means е:
   1. лаком алгоритъм
   2. оптимален алгоритъм
   3. алгоритъм използващ търсене в дълбочина с възврат
   4. глобално търсещ алгоритъм
2. Алгоритъмът k-means е:
   1. локално търсещ алгоритъм
   2. оптимален алгоритъм
   3. алгоритъм използващ търсене в дълбочина с възврат
   4. глобално търсещ алгоритъм
3. Кое от следните множества от примери съдържа шум
   1. (i)
   2. (ii)
   3. (iii)



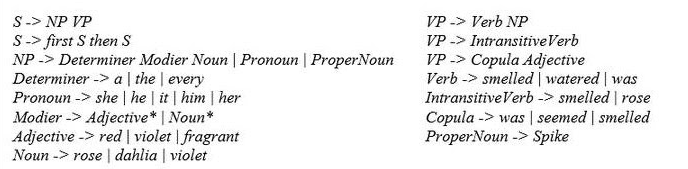
1. Какво е частично нареден план?

<http://fmi.wikidot.com/ai05> ctrl+f частично наредени планове lesno e

1. Кое от следните множества от примери е линейно разделимо
   1. (i)
   2. (ii)
   3. (iii)



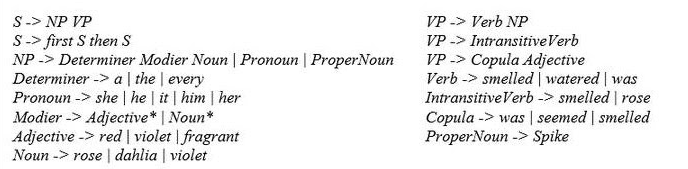
1. Кое от изброените клъстерни алгоритми изгражда йерархичен модел отгоре надолу
   1. k-means++
   2. йерархичен k-means
   3. агломеративен
   4. нито един от изброените
2. PCA (Principal component analysis)
   1. Решава класификационна задача
      1. примери
   2. Решава регресионна задача
   3. Свежда примерите до представителна извадка с по-малък размер
3. Ученето основано на примери (instance based learning) (kNN) може да бъде определено като
   * 1. мързеливо учене
     2. учене по аналогия
   1. И двете твърдения са верни
   2. Твърдение 1 е вярно, а 2 е невярно
   3. Твърдение 2 е вярно, а 1 невярно
   4. И двете твърдения не са верни
4. За k-means са дадени 2 твърдения:
   * 1. трябва да се направят няколко произволни рестартирания
     2. строи йерархичен класификатор
   1. И двете твърдения са верни
   2. Твърдение 1 е вярно, а 2 е невярно
   3. Твърдение 2 е вярно, а 1 невярно
   4. И двете твърдения не са верни
5. Дадена е следната контекстно-свободна граматика, където X\* означава 0 или повече срещания - Това не е за НАС - това е от NLP



Изречението the red red rose rose rose може да бъде генерирано от дадената граматик а. Посочете дали твърдението е грешно или не

* 1. Истина
  2. Лъжа

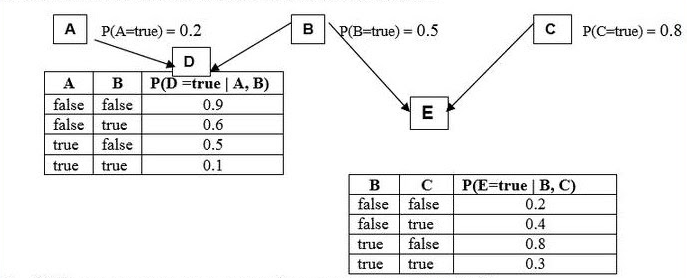
1. Дадена е следната контекстно-свободна граматика, където X\* означава 0 или повече срещания Това не е за НАС - това е от NLP



Изречението she was a violet violet violet може да бъде генерирано от дадената граматика. Посочете дали твърдението е грешно или не

* 1. Истина
  2. Лъжа

1. Посочете кои алгоритми спадат към учене без учител (Unsupervised learning)
   1. kMeans
   2. C4.5
   3. kNN
   4. EM-алгоритъм
   5. DBSCAN
   6. ID3
2. Дадена е Бейсова мрежа, където А–Е са булеви променливи



Каква е вероятността всичките булеви променливи да са true

* 1. 0.0024
  2. 0.00144
  3. 0.03
  4. 0.08

1. Ако Температура е условно независима от Главоболие при дадена диагноза Грип, то вярно или невярно е всяко едно от твърденията
   * 1. P(Главоболие , Температура | Грип) = P(Главоболие | Грип)P(Температура | Грип)
     2. P(Температура | Главоболие, Грип) = P(Температура | Грип)
   1. вярно, вярно
   2. вярно, невярно
   3. невярно, вярно
   4. невярно, невярно
2. Ако Температура е условно независима от Главоболие при дадена диагноза Грип, то вярно или невярно е всяко едно от твърденията
   * 1. P(Главоболие | Температура, Грип) = P(Главоболие | Грип)
     2. P(Грип | Главоболие, Температура) = P(Грип)P(Главоболие | Грип)P(Температура | Грип) / P(Температура)P(Грип)
   1. вярно, вярно
   2. вярно, невярно
   3. невярно, вярно
   4. невярно, невярно
3. Ако Температура е условно независима от Главоболие при дадена диагноза Грип, то вярно или невярно е всяко едно от твърденията
   * 1. P(Температура|Главоболие, Грип) = P(Температура|Грип) P(Температура|Главоболие)
     2. P(Температура, Главоболие |Грип) = P(T|Грип) P(Температура|Главоболие)
   1. вярно, вярно
   2. вярно, невярно
   3. невярно, вярно
   4. невярно, невярно
4. Индукция на дървото ID3 учи локален модел
   1. вярно
   2. невярно
5. Невронната мрежа учи локален модел
   1. истина
   2. лъжа
6. При обучение на логистична регресия се използва метод на градиентно изкачване
   1. истина
   2. лъжа
7. Може ли “изключващо или” xor да се реализира с невронна мрежа с едно скрито ниво?
   1. да
   2. не
8. Кой от изброените клъстерни алгоритми изгражда нейерархичен модел?
   1. Агломеративен
   2. bisecting k-means
   3. нито един от изброените
   4. k-means
9. Правото (от началното състояние към целта (Progressive)) и Обратното (от целта към началното състояние (Regressive)) търсене се използват само за строене на изцяло (totally) наредени планове.

а. Истина

б. Лъжа

1. Логистичната регресия предполага условна независимост на атрибутите/характеристиките
   1. Истина
   2. Лъжа
2. Дадени са съвместните вероятности

Positive

circle square

red 0.2 0.02

blue 0.02 0.01

Negative

circle square

red 0.05 0.3

blue 0.2 0.2

Пресметнете

P(positive | red ^ circle)

* 1. 0.25
  2. 0.80
  3. 0.57
  4. 0.20

1. За k-means алгоритъма са дадени следните 2 твърдения:
   * 1. Трябва да се направят няколко произволни рестартирания
     2. Строи йерархичен класификатор
   1. Твърдение 1 е вярно, твърдение 2 невярно
   2. Двете твърдения са неверни
   3. Твърдение 1 е невярно, твърдение 2 вярно
   4. Твърдения 1 и 2 са верни