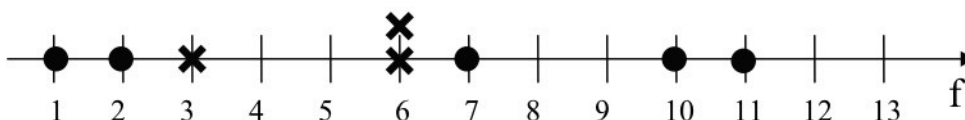


Zad. 1. Odpowiedz na pytania dotyczące poniższej macierzy pomyłek.

		odp. klasyfikatora	
		T	F
odp. oczekiwana	T	35	10
	F	15	40

- a) Ile sumarycznie odpowiedzi klasyfikatora było poprawnych?
- b) Ile sumarycznie odpowiedzi klasyfikatora było niepoprawnych?
- c) Ile wynosi trafność klasyfikacji?

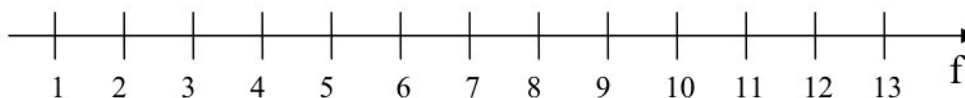
Zad. 2. Poniższy rysunek prezentuje zbiór danych uczących opisanych pojedynczym atrybutem warunkowym f . Atrybut decyzyjny przyjmuje jedną z dwóch wartości (klas decyzyjnych) – X i O (●). *Uwaga: przykłady uczące oznaczone są poniżej jako krzyżyki i kółka. Na osi f są tylko wartości atrybutu.*



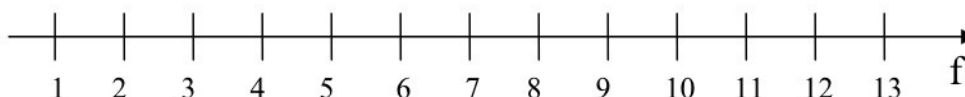
a) Powyższe dane zostały użyte do nauczania klasyfikatora **1-NN**, a następnie do klasyfikacji nieznanymi wcześniej danych. Jakie decyzje zostaną podjęte przez klasyfikator dla danych o wartościach atrybutu f równych:

$f_1=1$: _____, $f_2=2.4$: _____, $f_3=6.5$: _____, $f_4=9$: _____, $f_5=12$: _____

b) Zaznacz na poniższej osi przedziały wartości cechy f , dla których nowy obiekt zostanie zaklasyfikowany do klasy O (●).



c) Dane zostały również użyte do nauczania klasyfikatora **5-NN**. Zaznacz na osi przedziały wartości cechy f , przy których nowy obiekt zostanie zaklasyfikowany do klasy O (●).



Do wykonania poniższych zadań wykorzystaj program WEKA. Zapoznaj się wcześniej z krótką instrukcją obsługi zamieszczoną na stronie (plik *instrukcja_weka.pdf*).

Zad. 3. Poniższa tabela przedstawia dane uczące dla problemu klasyfikacji binarnej (2 klasy decyzyjne) o dwóch atrybutach numerycznych **A** i **B**. Dane te pochodzą z pliku *zad3.arff*. Korzystając z programu WEKA uzupełnij tabelę decyzjami zwracanymi przez klasyfikatory dla poszczególnych przypadków testowych i podaj skuteczności (trafności) poszczególnych klasyfikatorów.

Konfiguracja: By WEKA wypisywała odpowiedź klasyfikatora dla poszczególnych przypadków (predykcje), należy w zakładce '*Classify*' znaleźć panel '*Test options*', kliknąć w przycisk '*More options...*' i ustawić opcję '*Output predictions*' na PlainText. W tym samym panelu powinno być zaznaczone '*Use training set*' – dzięki temu skuteczność klasyfikatora będziemy badać na tych samych danych, na których się uczył (w ogólności jest to zła praktyka i chcemy zobaczyć, dlaczego).

Dane uczące			Odpowiedzi klasyfikatora			
A	B	Decyzja	1-NN	3-NN	11-NN	15-NN
1	5	0				
2	6	0				
2	7	1				
3	7	0				
3	8	1				
4	8	0				
5	1	1				
5	9	0				
6	2	1				
7	2	0				
7	3	1				
8	3	0				
8	4	1				
9	5	1				
10	6	1				
		Skuteczność				

a) Co się dzieje, gdy sąsiedztwo wynosi 1 (1-NN)?

b) Co się dzieje, gdy sąsiedztwo zawiera wszystkie przykłady uczące (15-NN)?

c) Podaj trafności klasyfikatorów k-NN dla k=1, 3, 11, 15 na zbiorze testowym wygenerowanym jako:

- podzbiór oryginalnego zbioru danych (percentage split, 33%)

- 15-krotna walidacja krzyżowa (cross-validation, 15 folds)

Zad. 4. Używając oprogramowania WEKA zbuduj klasyfikator k-NN dla problemu diagnozowania białaczki, dla którego dane umieszczone są w pliku leukemia.csv. Przy jakim k trafność klasyfikatora (oceniana na podstawie 10-krotnej walidacji krzyżowej) jest największa? Ile ona wynosi?