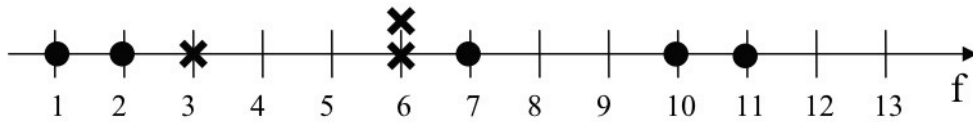
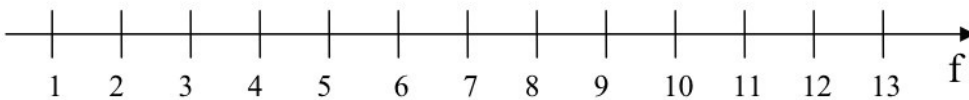


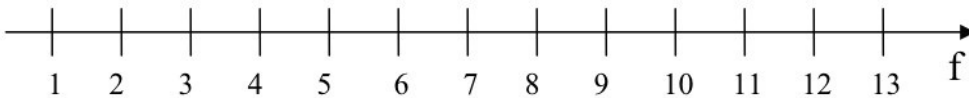
1. Poniższy rysunek prezentuje zbiór danych uczących opisanych pojedynczym atrybutem warunkowym f . Atrybut decyzyjny przyjmuje jedną z dwóch wartości (klas decyzyjnych) – X i O.



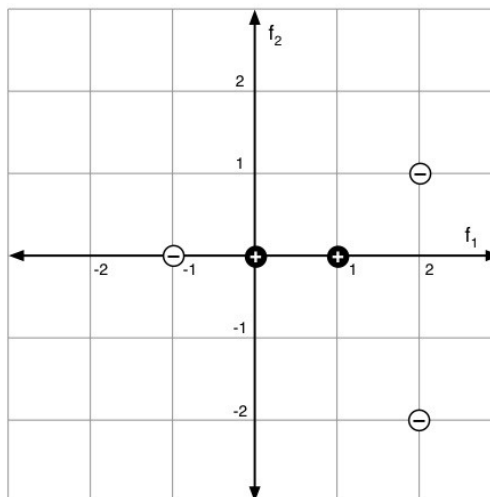
- 1.1. Powyższe dane użyte są do nauczania klasyfikatora minimalnoodległościowego **1-NN**, a następnie klasyfikacji nieznanymi wcześniej danych. Zaznacz na osi przedziały wartości cechy f , przy których nowy obiekt zostanie zaklasyfikowany do klasy O.



- 1.2. Powyższe dane użyte są do nauczania klasyfikatora minimalnoodległościowego **5-NN**, a następnie klasyfikacji nieznanymi wcześniej danych. Zaznacz na osi przedziały wartości cechy f , przy których nowy obiekt zostanie zaklasyfikowany do klasy O.



2. Na poniższym wykresie zaznaczone zostały dane uczące. Naszkicuj granice decyzyjne, których używałby klasyfikator minimalnoodległościowy **1-NN** do klasyfikacji nowych przykładów. Do której klasy decyzyjnej zaklasyfikowany zostałby nowy obiekt o atrybutach $f_1=1$ i $f_2=-1.01$ przy użyciu reguły **1-NN**, a do jakiej używając **3-NN**?



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

4. Dokonaj analizy zbioru danych uczących znajdujących się w pliku *komputery.arff*.
 - 4.1. Jakie są liczebności przykładów należących do każdej z klas decyzyjnych?
 - 4.2. Jakie podziały obiektów generuje każdy z atrybutów warunkowych?
 - 4.3. Oblicz miarę entropii dla klasyfikacji wszystkich przykładów.
 - 4.4. Używając entropii warunkowej i miary przyrostu informacji określ który atrybut zostanie wybrany jako najkorzystniejszy do zbudowania testu w korzeniu drzewa.
 - 4.5. Przy użyciu programu WEKA przeprowadź proces budowy całego drzewa (algorytmem ID3). Naszkicuj poniżej drzewo wygenerowane przez program.

5. Używając programu WEKA zbuduj drzewo decyzyjne (algorytmem J48) dla problemu diagnozowania raka piersi, dla którego dane uczące umieszczono w pliku *breast-cancer.arff*.
 - 5.1. Porównaj drzewa wygenerowane z użyciem mechanizmu upraszczania i bez niego. Jakie są różnice w rozmiarach drzewa a jakie w skuteczności klasyfikacji? Weź pod uwagę skuteczność na zbiorze uczącym i zbiorze testowym (uzyskanym za pomocą podziału pierwotnego zbioru danych metodami *hold-out* i *k-fold cross validation*).
 - 5.2. Przeanalizuj działanie poszczególnych mechanizmów upraszczania drzewa – jak je oceniasz?