**WSI  
Zadanie 7  
Naiwny klasyfikator Bayes’a  
Jakub Cholewiński**

1. **Implementacja naiwnego klasyfikatora Bayes’a.**

Zadanie polegało na utworzeniu klasyfikatora do gry w Snake’a, tak samo jak w poprzednich zadaniach. Do stworzenia klasyfikatora Bayes’a użyłem takich samych atrybutów jak w poprzednich zadaniach tzn. czterech binarnych mówiących o tym czy w kratce obok znajduje się przeszkoda, oraz czterech binarnych wskazujących w którym kierunku znajduje się nagroda. Kolejnym zapożyczeniem z poprzednich zadań jest zbiór danych na podstawie których stworzyłem agenta do gry.

Przechodząc już do implementacji, można ją podzielić na dwa etapy. Najpierw na podstawie danych uczących liczymy prawdopodobieństwa, które następnie wykorzystujemy do predykcji. Predykcja odbywa się za każdym razem na podstawie wyznaczonych offline prawdopodobieństw.

1. **Zastosowanie naiwnego klasyfikatora Bayes’a do gry Snake.**

Na podstawie zapisanego kodu, przeszedłem do takiego dobrania parametrów, aby agent zdobywał minimum 10 punktów w trakcie jednej rozgrywki. W trakcie trenowania, doszedłem do wniosku, że model jest bardzo wrażliwy na jakość danych uczących. Dołożenie lub odjęcie niewielkiej ilości danych, może powodować spore zmiany w decyzjach podejmowanych przez agenta. Na przykład przy wykorzystaniu 25 % utworzonych danych, model zdobywał średnio 10,5 punkta na rozgrywkę, natomiast wykorzystanie 35 % dostępnych danych, obniża tą średnią do 9. Dobieranie odpowiedniej ilości danych, było jedynym parametrem możliwym do sterowania w tym projekcie. Poniżej przedstawiam testy, dla różnych wielkości zbioru uczącego:

|  |  |  |  |
| --- | --- | --- | --- |
| Wielkość zbioru uczącego | Dokładność dla zbioru treningowego | Dokładność dla zbioru walidującego | Średnia zdobytych punktów na 100 gier |
| 1 | 0,8475 | 0,8410 | 9,46 |
| 0,75 | 0,8506 | 0,8778 | 9,34 |
| 0,5 | 0,8416 | 0,8524 | 9,94 |
| 0,25 | 0,8379 | 0,8071 | 10,69 |
| 0,15 | 0,8726 | 0,9152 | 9,73 |

W przypadku wybranej wielkości zbioru uczącego można powiedzieć, że występuje niedouczenie. Ciężko powiedzieć dlaczego akurat dla takiego zbioru treningowego, agent zachowuje się najlepiej, lecz może być to związane z jakością danych.