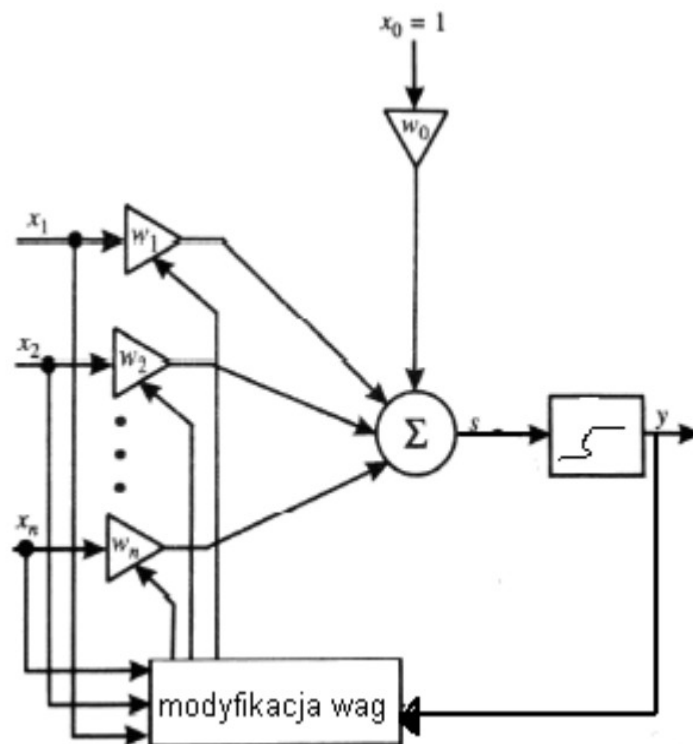


### Syntetyczny opis budowy sieci:

Model ten ma identyczną strukturę jak w przypadku modelu typu Adaline oraz neuronu sigmoidalnego, ale charakteryzuje się specyficzną metodą uczenia, znaną pod nazwą reguły Hebb'a. Reguła ta występuje z nauczycielem jak i bez nauczyciela. Hebb zauważył podczas badań działania komórek nerwowych, że połączenie pomiędzy dwiema komórkami jest wzmacniane, jeżeli w tym samym czasie obie komórki są aktywne.



Zaproponował on algorytm, zgodnie z którym modyfikację wag przeprowadza się następująco:

$$w_i(t+1) = w_i(t) + \eta * y * x_i$$

Oznaczenia:

- i-numer wagi neuronu,
- t-numer iteracji w epoce,
- y-sygnał wyjściowy neuronu,
- x-wartość wejściowa neuronu,
- $\eta$  - współczynnik uczenia (0,1).

W przypadku pojedynczego neuronu w trakcie uczenia będziemy modyfikować wartość wag proporcjonalnie zarówno do wartości sygnału podanego na i-te wejście, jak i sygnału wyjściowego  $y$  z uwzględnieniem współczynnika uczenia. Zauważmy, że w tym przypadku nie podajemy wzorcowej wartości wyjściowej, stosujemy więc tu metodę uczenia bez nauczyciela. Niewielka modyfikacja algorytmu:  $w_i(t+1) = w_i(t) + \eta * d * x_i$  prowadzi do drugiej metody uczenia neuronu Hebb'a: z nauczycielem (d-sygnał wzorcowy). Pewną wadą omawianego przez nas algorytmu jest to, iż wartości wag mogą wzrastać do dowolnie dużych liczb.

Dlatego też w praktyce stosuje się warianty metody uczenia z tzw. współczynnikiem zapominania, odpowiednio:

$$w_i(t+1)=w_i(t)*(1-\gamma)+\eta*y*x_i$$

$$w_i(t+1)=w_i(t)*(1-\gamma)+\eta*d*x_i$$

gdzie:

$\gamma$  – współczynnik zapominania

Sieć używana w danym zadaniu składała się z jednego neuronu Hebb'a z racji na najlepsze uzyskane wyniki.

Sieć była dla wartości:

```
learningRate=[0.00011, 0.1]
```

forgetRate=[0.1, 0.82]

liczba epok=100

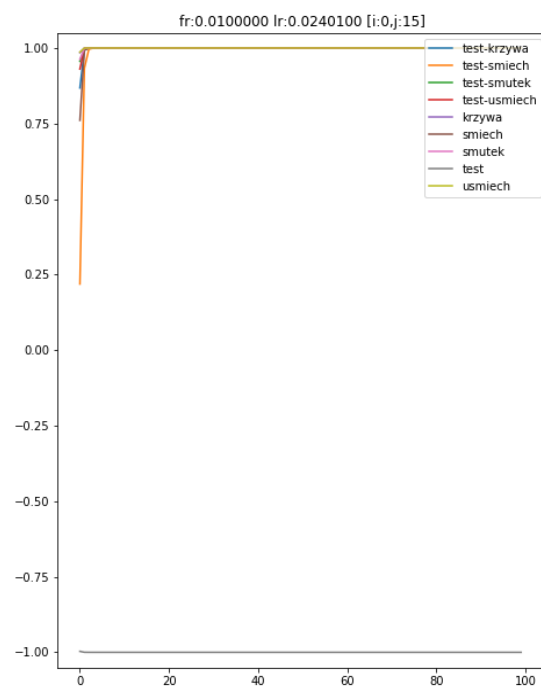
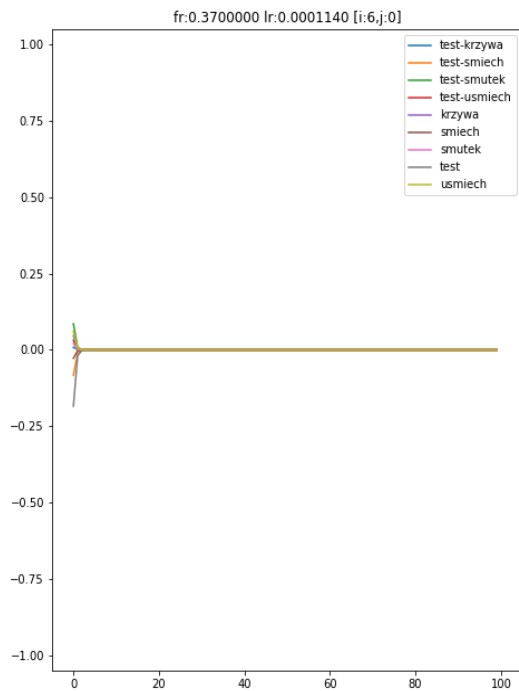
### Otrzymane wyniki:

Z racji na duży rozmiar otrzymanych wyników, dane są prezentowane w osobnych plikach

Wyniki obejmują 200 plików z danymi uzyskanymi dla tych samych danych uczących, to jest 4 rodzajów emotikon, wraz z jednym obrazkiem testowym, nieprzystającym do pozostałych, oraz 4 tych samych emotikon zaszumionych.

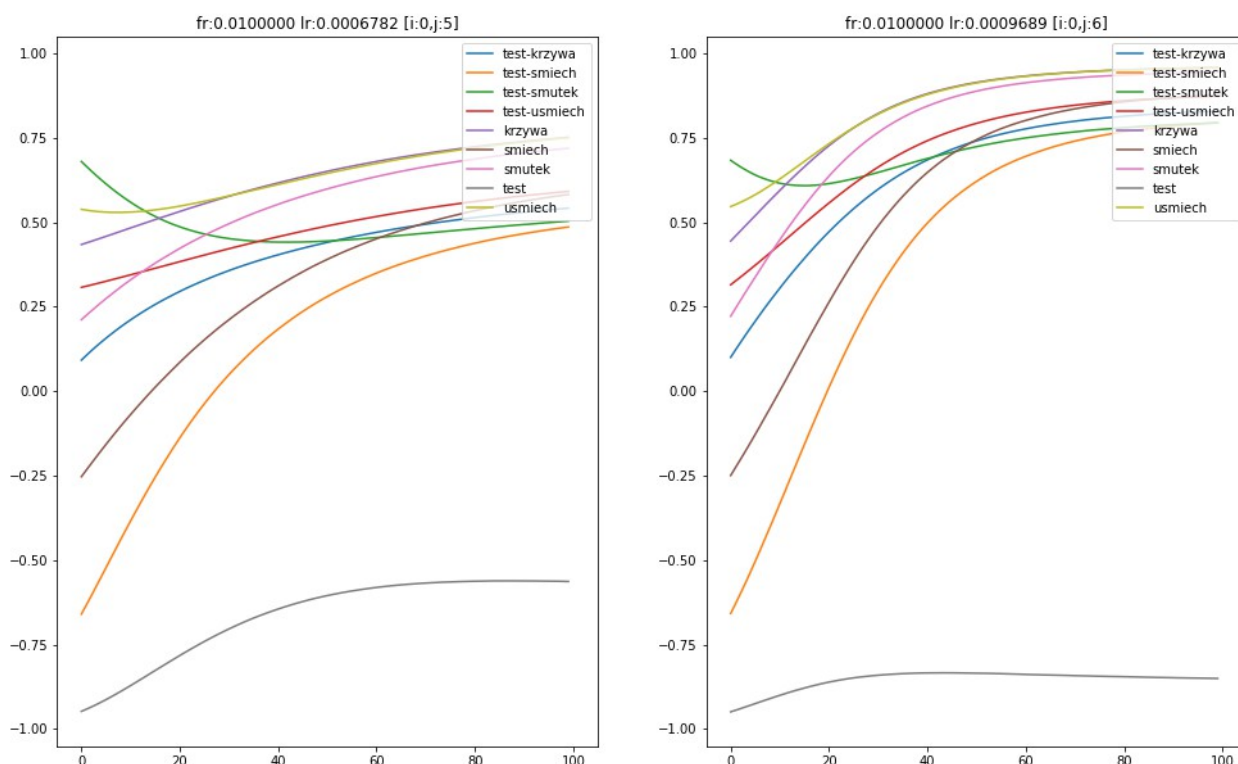
Z racji na bardzo dużo danych wyniki zestawienia są prezentowane na 200 różnych wykresach w jednym zestawieniu.

### Analiza wyników:



Widać że dla relatywnie dużych wartości współczynnika zapominania sieć dość szybko uczyła się w sposób, który dla każdej wartości wejściowej zwracał wartość jak najbardziej zbliżoną do 0.0. Oznacza to, że wagi były bardziej modyfikowane w kierunku 0.0, co jest zrozumiałe, gdyż każdorazowe pomnożenie wagi przez czynnik bliski zeru (1.0 – wsp. zapominania, przy wsp. zapominania  $\Rightarrow$  1.0) sprawia że wartość tej wagi czy to ujemna, czy dodatnia, dąży do 0.0. Widać również, że bardzo duże wartości współczynnika uczenia ostatecznie klasyfikowały różne dane wejściowe do dwóch podtypów o wartościach 1.0 i -1.0.

Najciekawsza część wyników znajduje się w środkowej części zestawienia. Gdzie można zauważyć zróżnicowany wpływ współczynnika uczenia i zapominania na uzyskiwane wyniki.



Dla bardzo małych współczynników zapominania (0.1) w stosunku do relatywnie małych współczynników uczenia (0.0006 – 0.001) uczenie przebiegało dobrze i można było wyróżnić więcej niż 2 końcowe podtypy danych. Wyniki uczenia nie wahały się podczas procesu uczenia – widać łagodne przejście od wag początkowych do wag końcowych.

W przypadku takiego uczenia można było ostatecznie wyróżnić 3-4 końcowe grupy danych:

1. krzywa, smutek, uśmiech
2. śmiech, test-uśmiech
3. test-smutek, test-śmiech
4. test

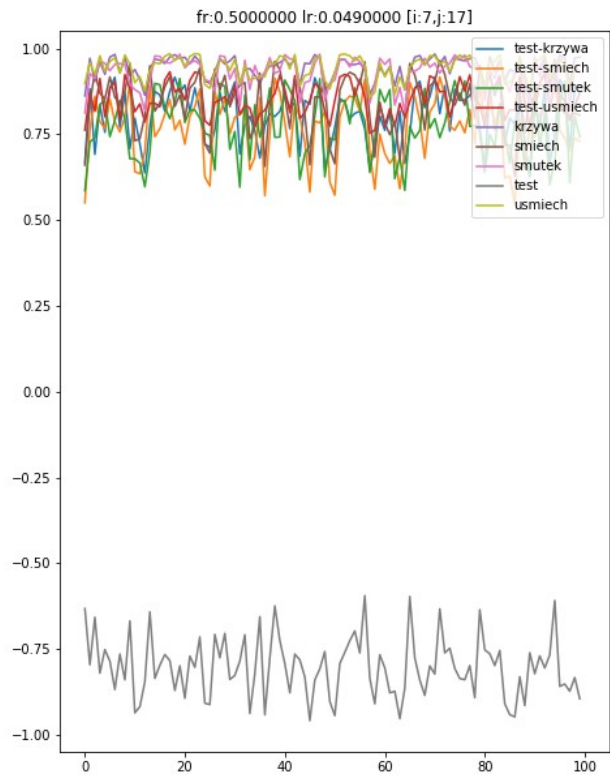
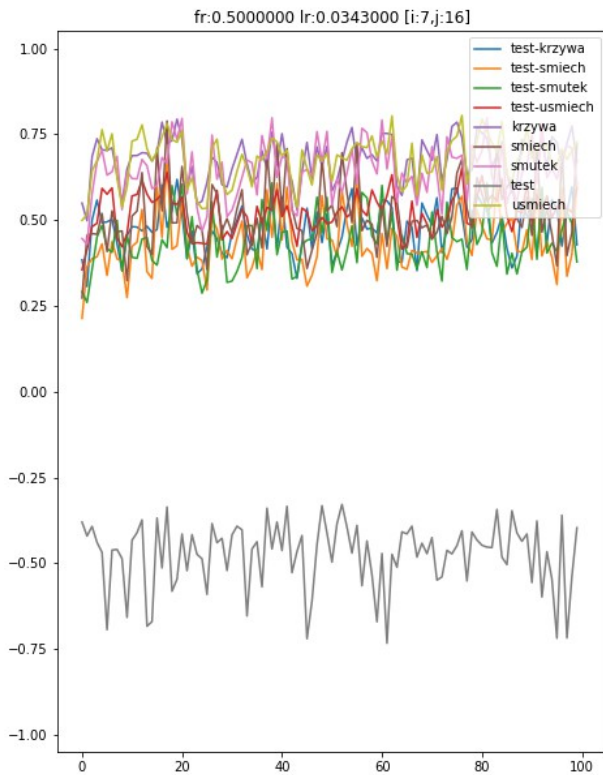
Minka „test-krzywa” kwalifikuje się między grupą 2, a 3.

Im bardziej zwiększamy współczynnik uczenia, tym bardziej uzyskiwane wyniki dążą do podziału na 2 grupy danych.

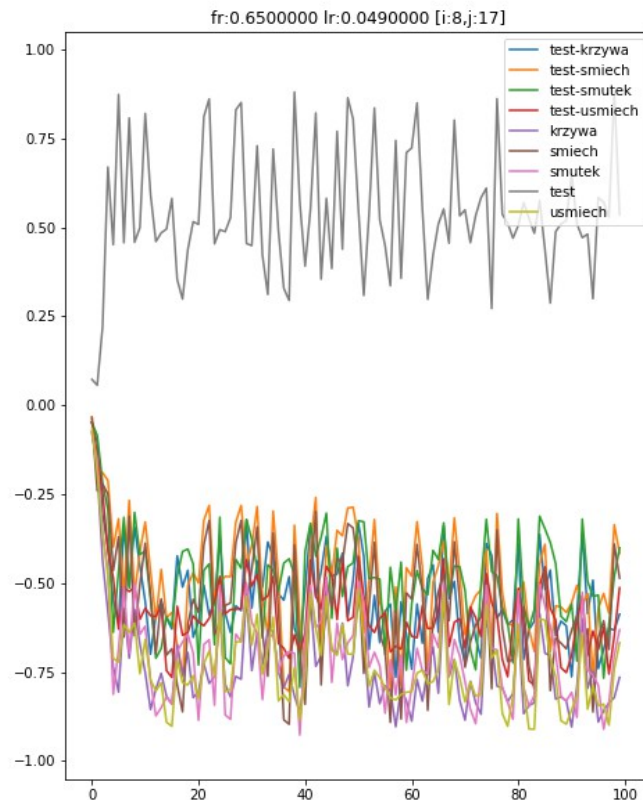
Jeżeli w miarę równomiernie będziemy zwiększać współczynnik zapominania oraz współczynnik uczenia, to będziemy uzyskiwali podział na podobną ilość grup. Jednakże zwiększymy szybkość

uczenia sieci i ostateczny podział będzie uzyskany szybciej.

W zależności też od tego z jaką dokładnością dobierzemy współczynniki – dane będą odpowiednio rozłożone na wykresie. Jeżeli będzie bardziej przeważać współczynnik zapominania – wtedy dane skupią się w środkowej części wykresu, będą bardziej zbliżone do 0, a poszczególne grupy powinny być łatwiej rozróżnialne. Jeżeli będzie przeważał współczynnik uczenia – wykresy poszczególnych „minek” będą się zbliżać do biegunów (-1.0, 1.0) przez co uzyskamy podział bardziej „binarny”.



Jeśli coraz bardziej zwiększamy współczynnik zapominania (przy stałym zwiększaniu współczynnika uczenia) – wtedy dane wyjściowe przez cały proces uczenia ulegają coraz to większym wahaniom. Jeśli będziemy zwiększać współczynnik do bardzo dużych wartości (0.8) wtedy ostateczne wyniki uczenia będą ulegały tak dużym wachaniom wartości z epoki na epokę, że nie będziemy w stanie odróżnić od siebie końcowych grup i będziemy zmuszeni potraktować dane wyjściowe jak dane binarne (wartości  $>0.0$  i  $<0.0$ )

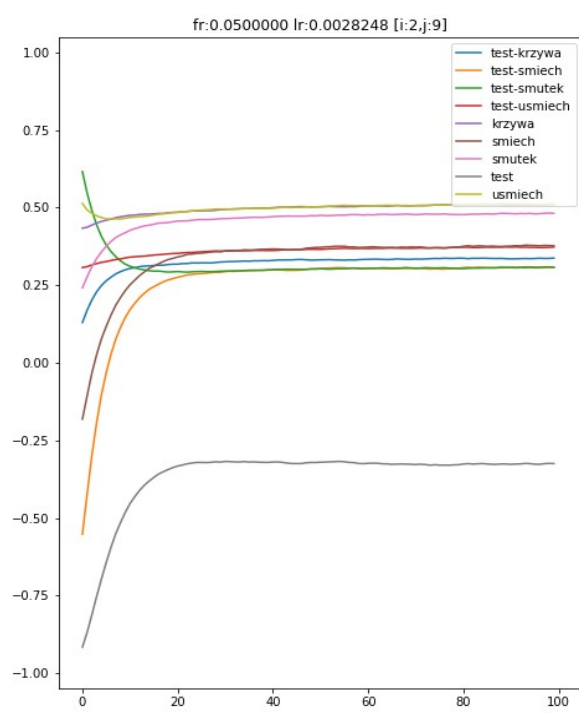
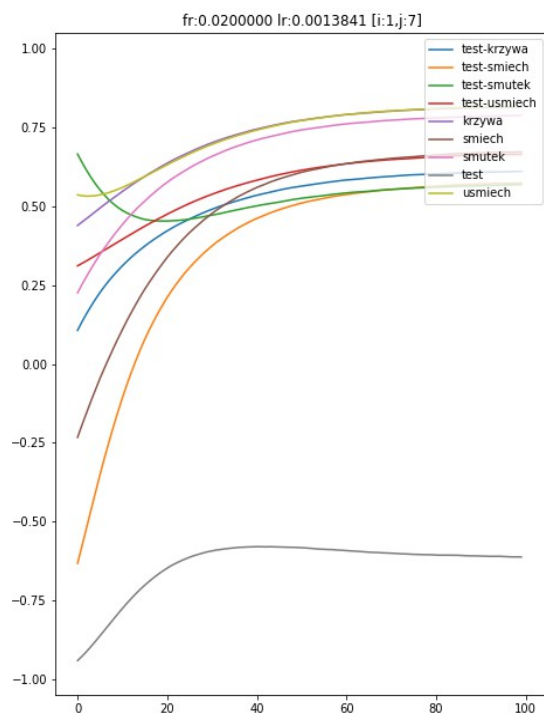


Ciekawym jest fakt, że dla pewnych szczególnych współczynników uczenia i zapominania, (przy dużych wartościach), sieć nauczyła się typów „na odwrót” niż inne badane sieci.

Najlepsze wyniki sieci zostały uzyskane dla wartości:

- $fr=0.02$ ,  $lr=0.0013841$
- $fr=0.05$ ,  $lr=0.0028248$

Stosunek  $fr/lr = [14.0, 18.0]$



**Wnioski:**

- Współczynnik zapominania nie może być ani zbyt duży, ani zbyt mały w stosunku do współczynnika uczenia, gdyż ostatecznie wartości zwracane przez zastosowaną sieć dążą do 0.0 , lub do dwóch wartości skrajnych: -1.0, 1.0. Stosunek wartości  $\text{fr}/\text{lr}$  powinien wynosić ok. 15 w celu uzyskania najlepszych wyników.
- Wartości bezwzględne współczynnika uczenia i zapominania nie powinny być zbyt duże, by uniknąć dużych wahań wartości wyjściowych z epoki na epokę.