

Raport wstępny

Sieć CNN do klasyfikacji obrazów znaków

Piotr Heinzelman
Wydział Elektryczny

1. Metodyka - Algorytm Sztucznej Inteligencji

W temacie zadania podany jest rodzaj sieci - sieć typu CNN.
Jest to sieć głęboka, składająca się z kilku współpracujących ze sobą warstw.

Znak (Obraz - plik graficzny o rozmiarze 250x250 pikseli w trybie szarości)
podawany jest w postaci macieży 2D (karty) W (250x250) liczb o zakresie 0-255 (stopień szarości) na wejście sieci neuronowej.

W pierwszym etapie wektor wejściowy jest poddawany filtrowaniu.
filtrowanie to jest realizowane przez operację zwaną splątaniem i polega na wyliczeniu ilorazu kolejnych próbek z wektorami filtrów. W wyniku filtrowania powstaje zbiór map wyjściowych H .

Filtrowanie może powodować pomijanie informacji przy krawędziach map - aby temu przeciwdziałać - należy zastosować warstwy Padding - powiększającą rozmiar wektora, uzupełnioną zerami, dzięki temu ważna informacja nie umyka przy operacji splątania.

Warstwa Poling - jest stosowana do zmniejszenia wymiarów wektora - zmniejsza one "rozdzielczość" wektora danych - jednak w taki sposób by informacje zostały zachowane przez zwiększenie "kontrastu" - jeśli użyć porównania z obrazem. Matematycznie - wybieramy największą (Max pooling) lub średnią (Avg pooling) wartość z kilku sąsiednich cel. Ewentualnie Lpooling.

Zmniejszone przefiltrowane dane podawane są do bloku dwuwarstwowego perceptronu ("Fully Connected") z funkcją aktywacji ReLU. Blok ten realizuje funkcję klasyfikatora obrazów. Uczenie sieci zgodnie z - WTA - zwycięzca bierze wszystko.

Wynik klasyfikacji jest podawany jest do ostatniej warstwy "soft-max" gdzie odpowiedź jest "konfekcjonowana" - czyli matematycznie przeliczana tak by suma wyników wynosiła 1, czyli wynik jest rozkładem gęstości prawdopodobieństwa każdej klasy.

Warstwa "soft-max" nie podlega uczeniu, nie ma ona parametrów - jednak trzeba uwzględnić jej wpływ na głębsze warstwy w procesie uczenia warstwy klasyfikatorskiej. Uczenie warstwy klasyfikatora polega na znalezieniu minimum funkcji błędów.

Pseudokod:

Uczenie sieci

```
Dopóki Średni_Błąd_Kwadratowy < zadana_wielkość_Błędu
{
    Dla każdego Obrazu z Listy_Danych_Treningowych
    {
        Zaprezentuj Obraz z listy Sieci (czyli  $X = \text{Obraz};$ )

        Znajdź odpowiedź sieci na wektor  $X$ ;

        Oblicz wielkość błędu odpowiedzi sieci dla Obrazu
        Oblicz i wprowadz poprawki wag w sieci klasyfikującej.

        Zapamiętaj błąd klasyfikacji dla Obrazu
    }
    Oblicz Średni_Błąd_Kwadratowy dla Epoki
    Zapamiętaj w bazie wielkość błędu epoki
}
```

Problem

Obraz - przekazujemy jako macierz 2D o wartościach od {0-255}.

Celem jest osiągnięcie minimum funkcji błędu klasyfikacji wektorów wejściowych i przypisania ich do 10 etykietowanych klas.

Weryfikacja

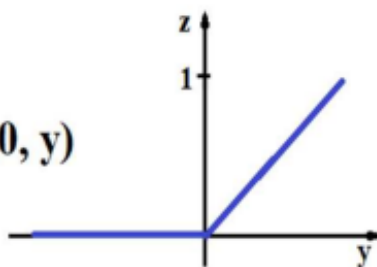
Weryfikacja - polega na sprawdzeniu odpowiedzi na obrazy których sieć “nie widziała” wcześniej. i sprawdzeniu czy średni błąd nie przekracza założonego marginesu błędu.

Operacje

funkcja aktywacji neuronu ReLU:

ReLU

$$z(y) = \max(0, y)$$



Splot dla filtra 3x3

$$h_{13} = \sum_{i=1}^3 \sum_{j=1}^3 w_{i,j} x_{i,j+2} + b$$

uczenie sieci klasyfikującej

minimalizacja błędu przez iteracyjne wprowadzanie poprawek wag dla WTA

$$\mathbf{w}_l(t+1) = \mathbf{w}_l(t) + \eta_l [\mathbf{x}(t) - \mathbf{w}_l(t)].$$

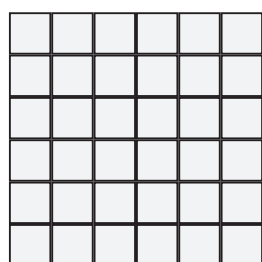
$$\mathbf{w}_k(t+1) = \mathbf{w}_l(t) + \eta_k G(k, l, \mathbf{x}(t)) [\mathbf{x}(t) - \mathbf{w}_k(t)].$$

$$G(k, l, \mathbf{x}(t)) = \begin{cases} 0 & \text{gd}y \quad k \neq l \\ +1 & \text{gd}y \quad k = l \end{cases}$$

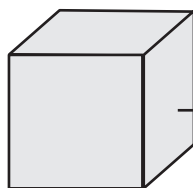
Soft-Max

$$s_k = \frac{\exp(z_k)}{\sum_{j=1}^K \exp(z_j)}, \text{ gdzie}$$

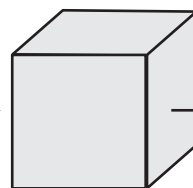
$$z_k = (y_k) = \left(\sum_{i=1}^N w_{ki} x_i + w_{k0} \right)$$



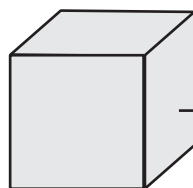
wejściowa
karta 2D



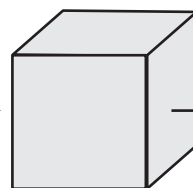
Filtr



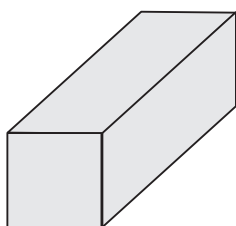
Max-Pool



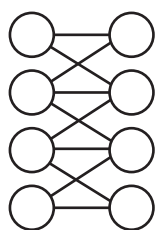
Filtry



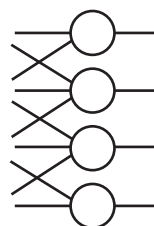
Max-Pool



Filtry



klasyfikacja
FC 2 warstwy



soft-max
wyjście



sprawdzenie
wyniku,
wyliczenie błędu