

# Dalszy rozwój algorytmu uczenia maszynowego dla ciągłej klasyfikacji stanu CP bozonu Higgsa w rozpadzie $H \rightarrow \tau\tau$

Jacek Kurek

January 17, 2020

## 1 Wstęp

Notka opisuje działania podjęte podczas pracy nad rozwojem i poprawa wydajności opartych na sieciach neuronowych algorytmów wyznaczania stanu CP bozonu Higgsa.

W trakcie pracy przy projekcie zajmowałem się zarówno poprawianiem technicznej strony kodu, a więc refactoringiem, usuwaniem powtórzeń czy dodaniem testów jednostkowych, jak i rozwijaniem poszczególnych metod podejścia do problemu.

Kod projektu implementuje różnorodne podejścia do klasyfikacji zdarzeń obserwowanych, w trakcie pracy w celu porównywania poszczególnych metod w jednolity sposób korzystałem z metryk (dających się określić w każdym przypadku) dotyczących odległości przewidywanego kata mieszania od właściwego dla danego przypadku, nawet jeżeli metoda nie używa tego kata explicite w funkcji kosztu.

## 2 Regresja do kata mieszania $\phi_{CP}$

Pierwsza zbadana metoda była regresja do kata mieszania - użyta funkcja kosztu był błąd średniokwadratowy. Metoda posiadała poważny problem który zaobserwować można na załączonym wykresie przedstawiającym histogram wygenerowanych zdarzeń, oraz histogram zdarzeń przewidzianych przez sieć:

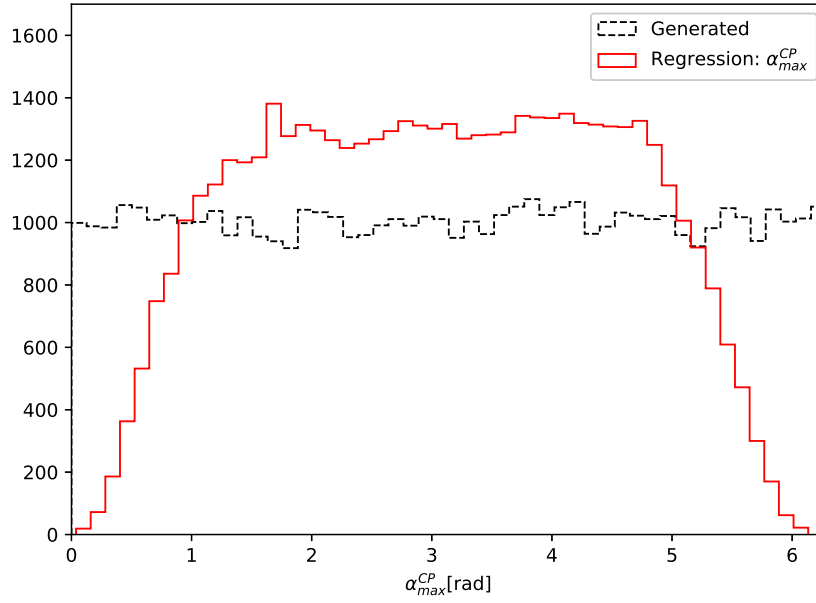


Figure 1: Wartość kąta mieszania dla wygenerowanych i przewidzianych zdarzeń bez użycia metryki okręgu

Jak można zaobserwować, sieć w znaczącym stopniu preferuje pośrednie wartości kąta, nie przewiduje nigdy wartości skrajnych.

Aby wyjaśnić to zachowanie, należy zauważyć, że wykorzystany tutaj błąd średniokwadratowy nie bierze pod uwagę faktu, iż wartości  $2\pi$  oraz 0 są w istocie tymi samymi wartościami. W efekcie wprowadza to niepożądaną nieciągłość i ściąga wyniki ku środkowi. Należało więc wykorzystać metodę która weźmie pod uwagę topologię przestrzeni wyjściowej.

Do porównania odległości dwóch kątów użyto funkcji  $d(\phi, \varphi) = 1 - \cos(\phi - \varphi)$  która opisuje odległość euklidesowa dwóch punktów na okręgu jednostkowym o współrzędnych katowych  $\phi, \varphi$ . Efekt można zaobserwować na następującym wykresie.

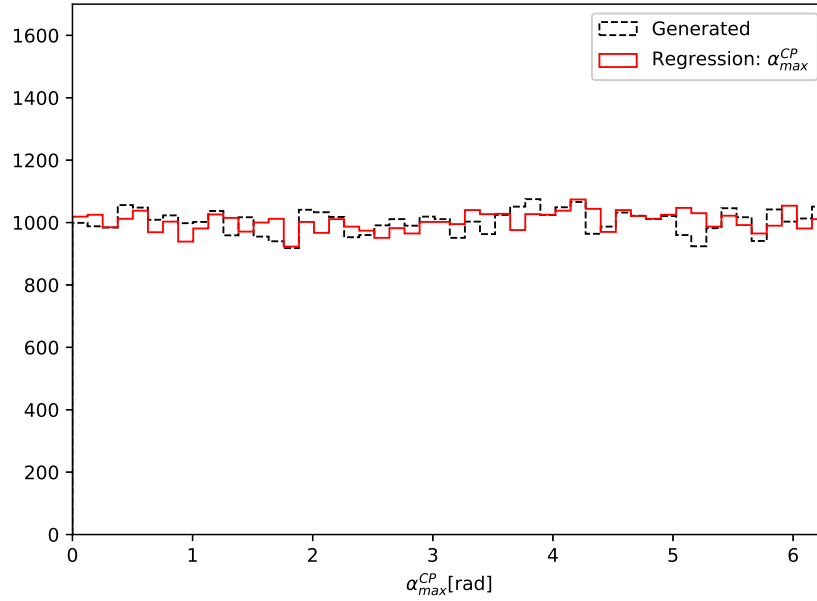


Figure 2: Wartość kąta mieszania dla wygenerowanych i przewidzianych zdarzeń przy użyciu metryki okręgu

Wraz z użyciem tego usprawnienia, poprawiły się metryki opisujące poprawność przewidywań realizowanych przez sieć neuronową:

	Przed usprawnieniem	Po usprawnieniu
Średni błąd przewidywania (rad)	0.574	0.435
Odchylenie standardowe (rad)	0.904	0.692

Table 1: Porównanie rezultatów przed i po zastosowaniu poprawek

### 3 Regresja do funkcji wagowej

Przypominając, dla pojedynczego eventu można ustalić funkcję  $w : [0; 2\pi] \rightarrow R$  postaci

$$w(\phi) = A + B \cos \phi + C \sin \phi$$

zaś wartość  $\phi_{CP}$  jest połową wartości kąta  $\phi$  dla której funkcja osiąga maksimum.

W tej metodzie na podstawie danych z eventu, przewidujemy zdyskretyzowany wykres funkcji  $w$  dla pewnej ustalonej liczby klas  $c$ . Jako funkcje kosztu używamy różnicy średniokwadratowej.

Metoda nie uczyła się poprawnie, generując wyniki charakteryzujące się wartościami parametrów  $B, C$  bliskimi zeru:

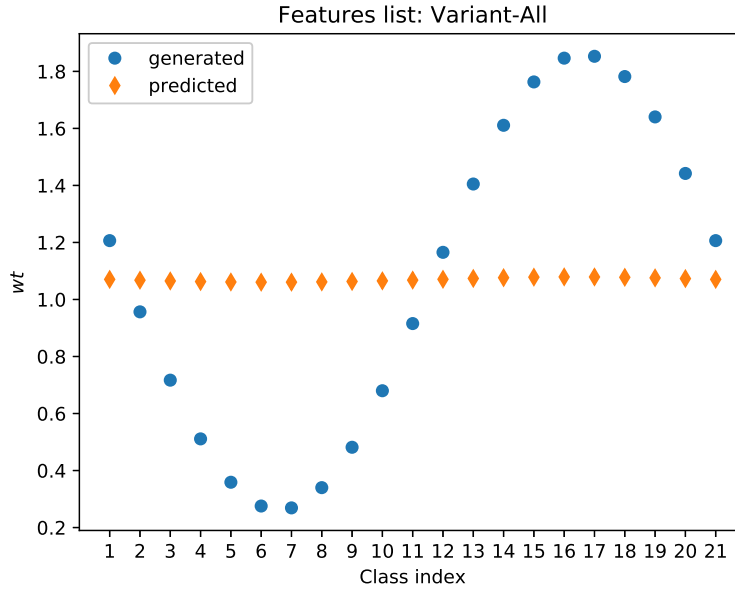


Figure 3: Przykładowy wykres przewidywanej funkcji wagowej względem oczekiwanej

Uwagę zwracał jednak fakt, że parametr  $A$  - wartość średnia funkcji, był dobrze ustalany. Na podstawie tej obserwacji, zaproponowana została metoda dostępna w kodzie przy użyciu flagi "delete-c", w której dane trenowane są po normalizacji:

$$W(\phi) = w(\phi) - \frac{\int_0^{2\pi} w(\varphi) d\varphi}{2\pi}$$

W efekcie, w istotnym stopniu poprawiły się zdolności predykcyjne. Poniższa tabela podsumowuje osiągnięte dokładności w zależności od użycia dropoutu i flagi "delete-c"

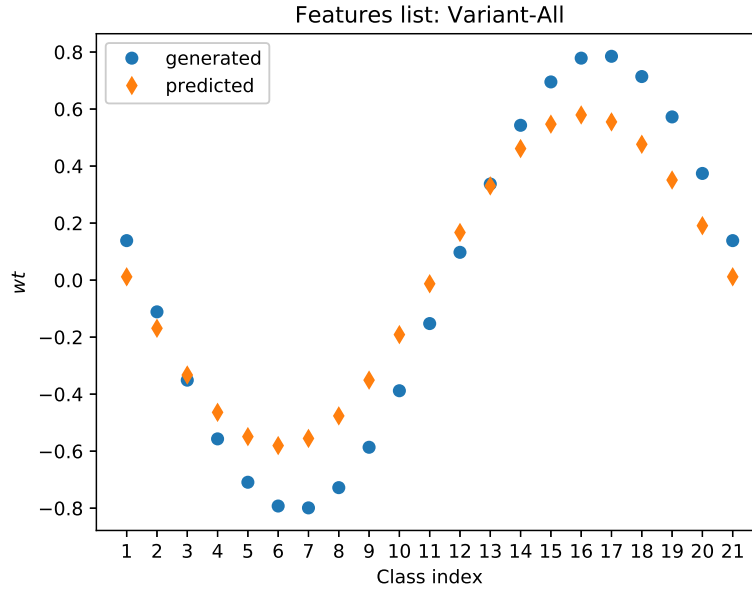


Figure 4: Przykładowy wykres przewidywanej funkcji wagowej względem oczekiwanej przy użyciu flagi `-delete-c`

	-delete-c	bez -delete-c
Dropout 20	0.548	1.531
Dropout 0	0.233	0.243

Table 2: Średni błąd przewidywania dla przypadków (rad)

	-delete-c	bez -delete-c
Dropout 20	0.810	1.778
Dropout 0	0.420	0.440

Table 3: Odchylenie standardowe dla przypadków (rad)

## 4 Wyznaczanie parametrów funkcji wagowej poprzez dyskretyzacje

Funkcja wagowa postaci

$$w(\phi) = A + B \cos \phi + C \sin \phi$$

może być wyznaczana poprzez dyskretyzację przestrzeni możliwych parametrów  $A, B, C$ .

W tym podejściu, używamy trzech instancji sieci, każda uczy się jednego z parametrów. Problem który należało rozwiązać, związany był z rozkładem parametrów przewidywanym przez sieć. Poniższy wykres pokazuje rozkład parametru  $A$  wyznaczony przez sieć, wraz z rozkładem tegoż parametru zawartym w danych wejściowych

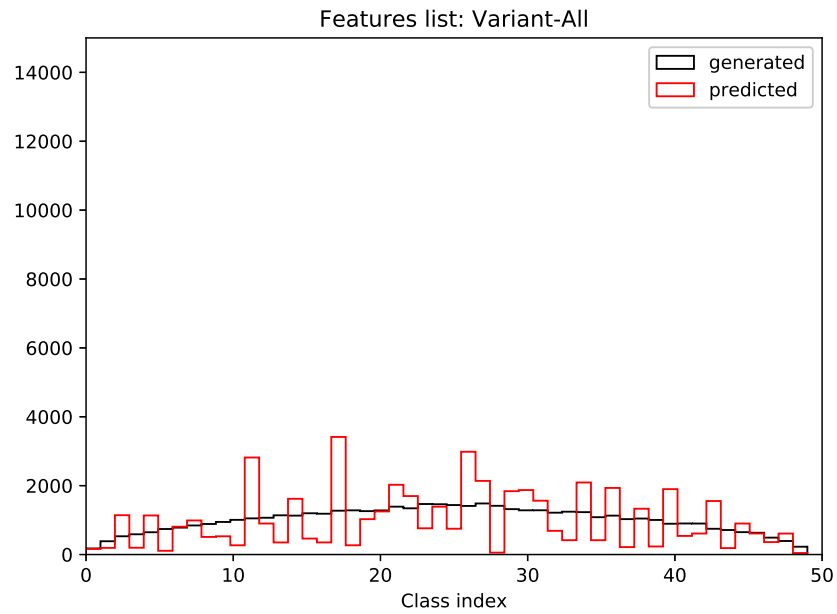


Figure 5: Rozkład parametru  $A$

Problem staje się jeszcze lepiej widoczny, jeśli spojrzymy na rozkład błędu:

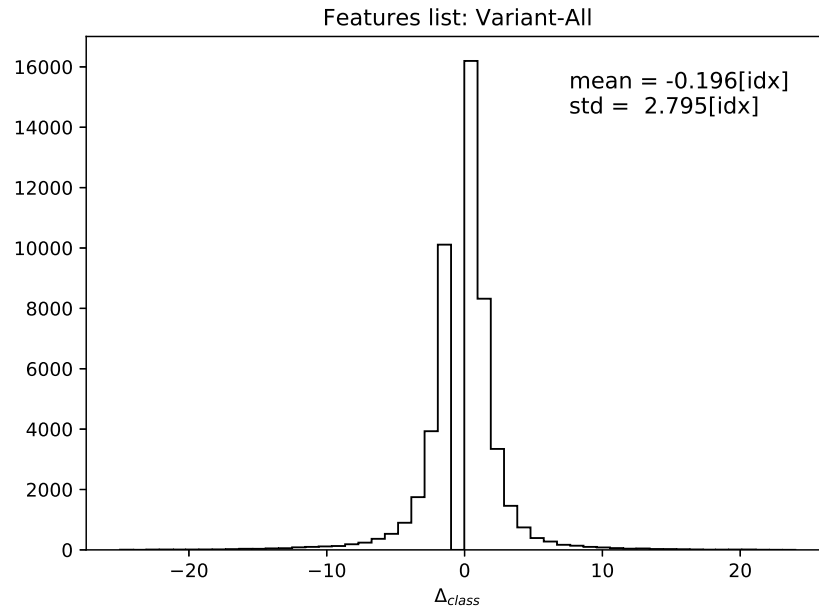


Figure 6: Błąd przewidywania

Isotne i powtarzalne różnice w przewidywaniach i rozkładach generowanych przez sieć neuronową sugerują potencjalny błąd implementacyjny. W ramach podjętych działań kod odpowiedzialny za generowanie danych w tym modelu został przejrany, niestety nie przyniosło to spodziewanych rezultatów. Rekomendowanym rozwiązaniem jest ponowna implementacja rozwiązania, włącznie z kodem odpowiedzialnym za dyskretyzację parametrów wejściowych.