

Úloha na lokální minima

Stanislav Novotný

December 1, 2021

1 Úvod

Zkoumáme využití neuronových sítí pro lineární regresi na jednoduchém příkladě fitování polynomu $p(x) = 2x^2$ na intervalu $[-1, 1]$ pomocí modelu $Chain(NaiveNPU(1, 1), Dense(1, 1))$. Z teorie víme, že lokálního minima nabudeme pro nastavení

$$W^r = 2, W^i = 0, A = 2, b = 0 \quad (1)$$

$$W^r = 0, W^i = 0, A = 0, b = \frac{2}{3} \quad (2)$$

První příklad představuje správný fit, zatím co druhá možnost je chybná.

Cílem je osvětlit jak moc palčivý tento problém je.

2 Experiment

V následujícím experimentu odhalíme, jak moc je tento problém zásadní. Podle počáteční inicializace parametrů modelu budeme zkoumat, různé případy.

1. Divergence
2. Konvergence do bodu (1)
3. Konvergence do bodu (2)
4. Konvergence do jiného bodu

Experiment rozložíme do dvou částí, v první budeme mít parametr b v každé počáteční inicializaci nastaven na hodnotu 0, ve druhé části budeme vždy začínat s hodnotou $b = \frac{2}{3}$. Určíme si intervaly, odkud budeme brát inicializační parametry W^r a A jako $W^r \in [-1, 3]$ a $A \in [-1, 3]$ v každém z těchto intervalů se budeme pohybovat s krokem 0.1. Parametr W^i bude vždy inicializován hodnotou $W^i = 0$. Ze všech takovýchto bodů spustíme iteraci za využití optimalizační techniky *ADAM* s krokem 0.05. Výsledek každého iterace uložíme jako prvek matice 41×41 , kterou následně vyhodnotíme podle následujících kritérií:

1. Je-li $\|\nabla W^r\| \geq \epsilon \vee \|\nabla A\| \geq \epsilon \vee \|\nabla b\| \geq \epsilon \rightarrow$ divergence
2. Je-li $|W^r - 2| < \epsilon \wedge |A - 2| < \epsilon \wedge |b| < \epsilon \rightarrow$ konvergence do bodu (1)
3. Je-li $|A| < \epsilon \wedge |b - \frac{2}{3}| < \epsilon \rightarrow$ konvergence do bodu (2)
4. Jinak \rightarrow konvergence do jiného bodu

Ve třetím bodě již nemusíme uvažovat podmínku $|W^r| < \epsilon$, jelikož zkoumaný polynom násobíme nulou díky podmínce $|A| < \epsilon$

Následující obrázky zobrazují výsledné matice tohoto experimentu. Barva na pozici (x, y) znázorňuje, jaké ze čtyř podmínek jsme vyhověli

- Žlutá \rightarrow divergence
- Zelená \rightarrow konvergence do bodu (1)
- Červená bodu (1) konvergence do bodu (2)
- Bílá konvergence do jiného bodu

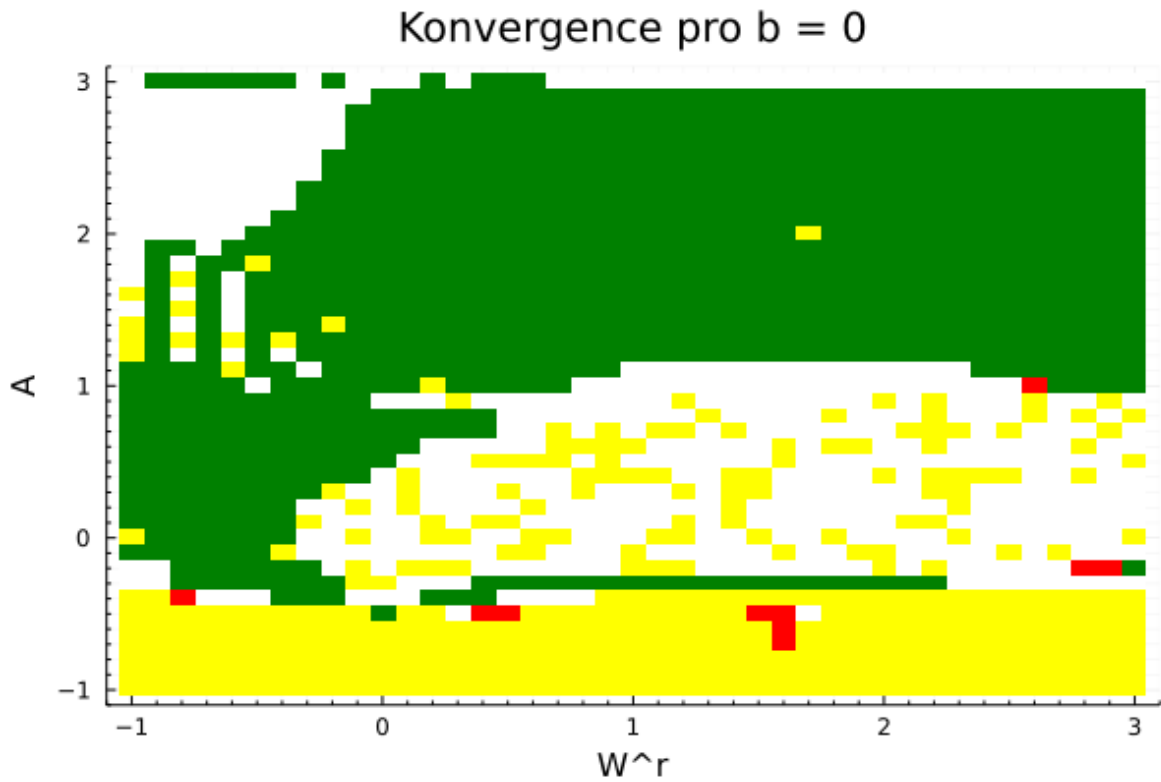


Figure 1: Zobrazení konvergence v závislosti na inicializaci parametrů W^r, A pro $b = 0$.

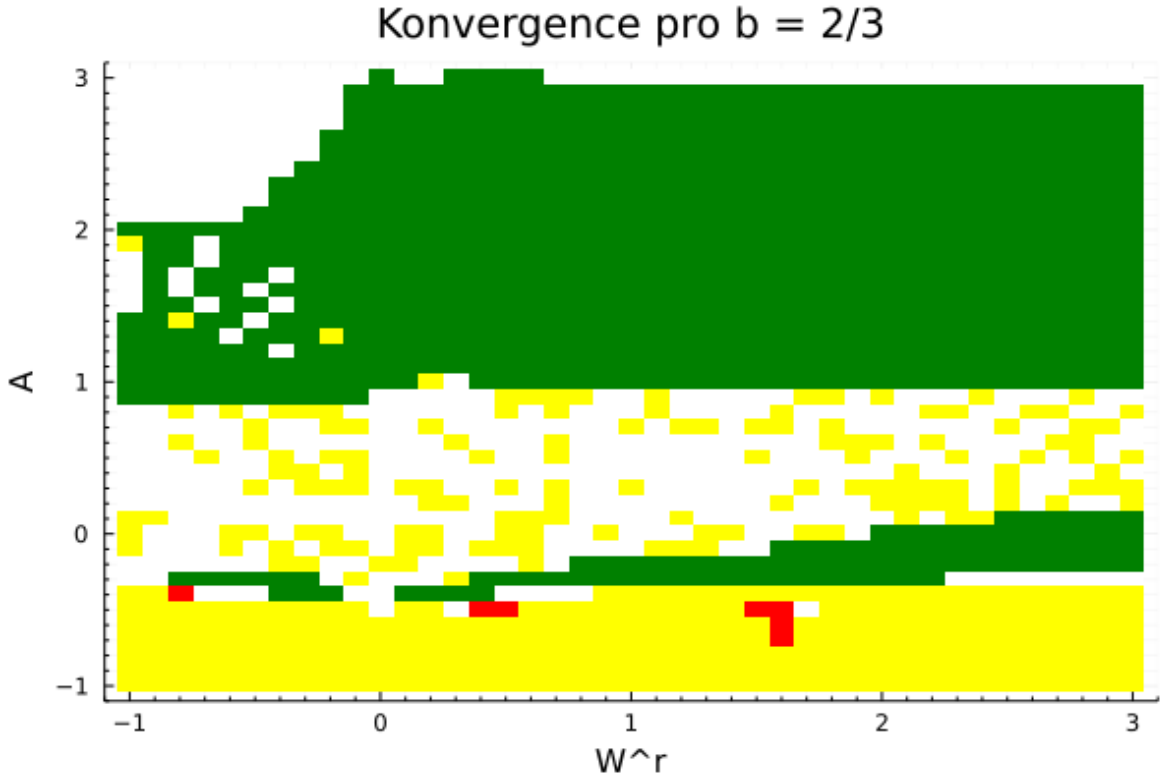


Figure 2: Zobrazení konvergence v závislosti na inicializaci parametrů W^r, A pro $b = \frac{2}{3}$.

Z obrázků pozorujeme, že správné konvergence se dosahovalo především v intervalu $A \in (1, 3)$, konvergence do (2) se objevovala pro $A \simeq -0.5$.

Následující tabulky představují numerickou interpretaci obrázků 1 a 2 pro hodnoty parametrů $b = 0$ a $b = \frac{2}{3}$. Matice na obou obrázcích jsou rozměru 41×41 , což představuje 1681 prvků.

Konvergence v závislosti na parametrech W^r, A pro $b = 0$		
Výsledek	Počet výskytů	Celkové zastoupení
Divergence	380	22.6%
Konvergence do (1)	836	49.73%
Konvergence do (2)	10	0.59%
Konvergence jinam	455	27.06%

Table 1: Tabulka konvergence v závislosti na inicializaci parametrů W^r, A pro $b = 0$

Konvergence v závislosti na parametrech W^r, A pro $b = \frac{2}{3}$		
Výsledek	Počet výskytů	Celkové zastoupení
Divergence	403	23.7%
Konvergence do (1)	835	49.67%
Konvergence do (2)	7	0.42%
Konvergence jinam	436	25.94%

Table 2: Tabulka konvergence v závislosti na inicializaci parametrů W^r, A pro $b = \frac{2}{3}$