1 **TASK 1.3** Pro jaké všechny hodnoty α algoritmus **diverguje**?

$\alpha_{\text{diverge}} \in$ $(-\infty; 0) \cup (0.2; +\infty)$ ✓

1 **TASK 1.4** Jaká je optimální rychlost učení α^*, při které je dosaženo **nejrychlejší konvergence** pro libovolnou inicializaci $\mathbf{w}^0$

$\alpha^* =$ 0.1 ✓

$1 - 10\alpha = 0$

$10\alpha = 1$; $\alpha^* = 0.1$

2 **TASK 1.5** **V kolika krocích** dosáhne gradientní sestup při použití α^* přesné optimální hodnoty (pokud nedosáhne nikdy uveďte ∞) a jaký je jeho convergence rate:

$N =$ 1 ✓

$\text{rate}(\alpha^*) =$ $\max\{|1-10\alpha|\} = 0$ ✓

1 **TASK 1.6** Jaká je optimální rychlost učení α^*, pro funkci $f(w_1, w_2) = 5w_1^2 + 2w_2^2 + 1$

$\alpha^* =$ $1/7$ ✓

$\frac{\partial f}{\partial w_1} = 10w_1$ $\frac{\partial f}{\partial w_2} = 4w_2$

$|1-10\alpha| = |1-4\alpha|$

$10\alpha - 1 = 1 - 4\alpha$

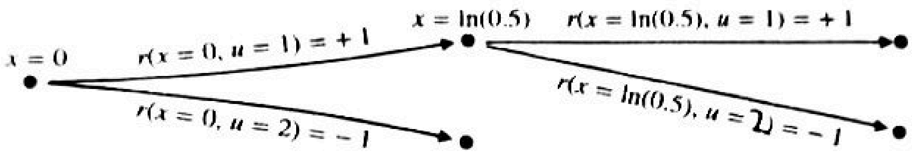
$14\alpha = 2$

$\alpha = 2/14 = 1/7$

2. Uvažujte stochastickou diskrétní policy, která vybírá akci $u \in \{1,2\}$ ve stavu $x \in \mathbb{R}$ takto:

$$\pi_\theta(u|x) = \begin{cases} \sigma(\theta x) & \text{if } u = 1 \\ 1 - \sigma(\theta x) & \text{if } u = 2 \end{cases}$$

kde $\theta = -1$ je skalární parameter. Prostor stavů, akcí a deterministických přechodů je dán obrázkem níže. Tato policy mapuje stav **x** na pravděpodobnostní rozdělení dvou možných akcí $u = 1$ nebo $u = 2$.



$\sigma(-\ln(0.5)) = 2/3$

Spočtěte následující pro $\gamma = 1$.

Tip: spočtěte si nejdříve hodnoty π_θ v jednotlivých uzlech s využitím $\sigma(\cancel{-\ln(0.5)}) = 1/3$:

$\sigma = \frac{1}{1+e^{-x}}$

0,5 **TASK 2.1** $Q^{\pi_\theta}(x = 0, u = 1) = 1 + \sigma(\theta x)\cdot 1 + (1-\sigma(\theta x))\cdot -1) = 1$ ×

$Q^{\pi_\theta}(x = 0, u = 2) = -1$ ✓

0,5 **TASK 2.2** $V^{\pi_\theta}(x = 0) = \sigma(\theta x)\cdot 1 + (1-\sigma(\theta x))\cdot(-1) = \frac{1}{2} - \frac{1}{2} = 0$ ×

$V^{\pi_\theta}(x = \ln(0.5)) = \sigma(\theta x)\cdot 1 + (1-\sigma(\theta x))\cdot(-1) = \frac{2}{3} - \frac{1}{3} = 1/3$ ✓

1 **TASK 2.3** $A^{\pi_\theta}(x = 0, u = 1) = Q^{\pi_\theta}(x=0, u=1) - V^{\pi_\theta}(x=0) = 1 - \cancel{0}^{4/6} = 1$ ✓

$A^{\pi_\theta}(x = 0, u = 2) = Q^{\pi_\theta}(x=0, u=2) - V^{\pi_\theta}(x=0) = -1 - \cancel{0}^{4/6} = -1$ ✓ } propagace chyby

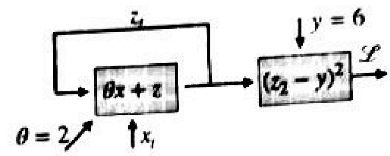
1 **TASK 2.4** Policy provede ve stavu $x = 0$ akci $u = 1$ (která byla vygenerována z jejího pravděpodobnostního rozdělení), a robot tím přejde do stavu $x = \ln(0.5)$. Spočtěte policy gradient a význam vypočteného čísla slovně interpretujte (1 krátká věta). ×

$$\frac{\partial \log \pi_\theta(u|x)}{\partial \theta}\bigg|_{\substack{x=0\\u=1}} \cdot A^{\pi_\theta}(x = 0, u = 1) = 0 \cdot A^{\pi_\theta}(x=0, u=1) = 0 \; ✓$$

$$\pi_\theta(u=1|x=0) = \frac{1}{1+e^{-\theta\cdot 0}} = \frac{1}{1+1} = \frac{1}{2}$$

$$\frac{\partial \log\left(\frac{1}{2}\right)}{\partial \theta} = 0$$

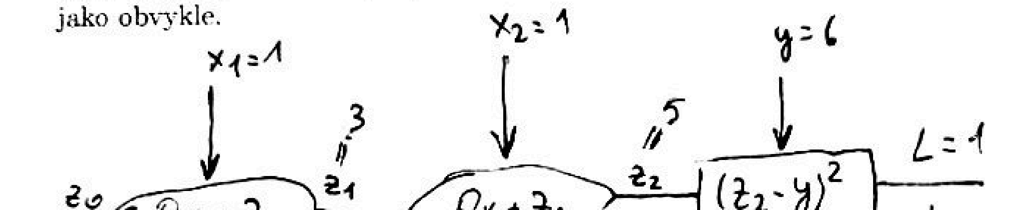
3. Uvažujeme rekurentní neuronovou síť znázorněnou na obrázku níže.



Síť je inicializována parametrem $\theta = 2$ a počátečním skrytým stavem $z_0 = 1$. Dostanete následující vstupní sekvenci: $x_1 = 1, x_2 = 1$. Ztráta se počítá pouze pro **poslední** skrytý stav z_2 jako L2 vzdálenost od ground truth hodnoty $y = 6$.

TASK 3.1 Spočtete gradient $\frac{\partial \mathcal{L}}{\partial \theta}$ ztráty $\mathcal{L}$ vzhledem k θ.

2 **Tip:** Rozviňte síť v čase, abyste získali dopřednou síť a proveďte backpropagation jako obvykle.



$\frac{\partial z_1}{\partial \theta} = \cancel{\theta}\ 1$ $\quad$ $\frac{\partial z_2}{\partial z_1} = \cancel{x}\ 1$ $\quad$ $\frac{\partial L}{\partial z_2} = 2z_2 - 2y = -2$ $\quad$ $\frac{\partial L}{\partial L} = 1$

$\frac{\partial z_2}{\partial \theta} = \cancel{x}\ 1$

Kontrola

$$L = ((\theta x_1 + z_0 + \theta x_2) - y)^2 = 1 = \theta^2 x_1^2 + 2\theta x_1 z_0 + 2\theta^2 x_1 x_2 + z_0^2 + 2\theta x_2 z_0 + \theta^2 x_2^2 - 2y(\theta x_1 + z_0 + \theta x_2)$$

$$\frac{\partial L}{\partial \theta} = 2\theta x_1^2 + 2x_1 z_0 + 4\theta x_1 x_2 + 2x_2 z_0 + 2\theta x_2^2 - 2y x_1 - 2y x_2 = -4$$

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial \theta} + \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial z_1} \cdot \frac{\partial z_1}{\partial \theta} = (-2 \cdot 2) + (-2 \cdot 2 \cdot 0) = -4 \checkmark$$

TASK 3.2 Proč je výsledný gradient kladný/nulový/záporný/nedefinovaný (jedna krátká věta)?

1 ✓

Protože máme θ zvětšit

$$\theta_{i+1} = \theta_i - \alpha \frac{\partial L}{\partial \theta}$$

$$\theta_1 = 2 + 4\alpha$$

VIR 2023
Exam test
Variant: A

Jméno: Jauheni Zviazdou

Body 15,5

1. Je dána funkce

$$f(w) = 5w^2 + 1.$$

Váhy této funkce $w \in \mathbb{R}$ budeme učit gradientním sestupem:

$$w^k = w^{k-1} - \alpha \left. \frac{\partial f^\top(w)}{\partial w} \right|_{w=w^{k-1}},$$

z iniciálního bodu w^0, kde α značí rychlost učení (learning rate).

1 **TASK 1.1** Pro jaké všechny hodnoty α algoritmus **konverguje** (alespoň pomalu)?

$\alpha_{\text{converge}} \in$ $(0, 0.2)$ ✓

$\frac{\partial f}{\partial w} = 10w$

$0 < 10\alpha < 2$

$0 < \alpha < 0.2$

1 **TASK 1.2** Pro jaké všechny hodnoty α algoritmus **osciluje**?

$\alpha_{\text{oscillate}} \in$ $\{0, 0.2\}$ ✓

2,5 4. Označte která z následujících tvrzení jsou správná (zaškrnutím čtverce):

- NE ☐ Zavedení skip-connection přes nějakou vrstvu zvětšuje její receptive field.
- NE ☐ Velikost receptive field závisí na velikosti gradientu.
- ANO ✓ ■ Velikost receptive field závisí na velikosti konvolučního jádra.
- NE ☐ Velikost receptive field závisí na batch_size (tj. velikosti batch který vstupuje do batch-norm vrstvy).
- NE ☐ Čím menší je receptive field tím lépe naučená síť generalizuje.
- ANO ✓ ■ Batch-norm omezuje explodování nebo mizení gradientu.
- × ■ Batch-norm zvyšuje pravděpodobnost saturace výstupních hodnot po sigmoid vrstvě.
- Asi NE ☐ Skip-connection zefektivňuje u segmentační sítě typu U-net rekonstrukci tvarů jednotlivých objektů.
- Asi NE ☐ Sémantická segmentace se sítí typu U-net umožňuje efektivní rekonstrukci konvexních objektů (tj. objektů ve tvaru písmene "U").

- Nahrazení **jedné** konvoluční vrstvy s jádrem 7x7, **třemi** konvolučními vrstvami s jádrem 3x3:
  - NE ☐ **sníží** receptive field ve výstupní příznakové masce.
  - ANO ✓ ■ **zachová** receptive field ve výstupní příznakové masce.
  - NE ☐ **zvýší** receptive field ve výstupní příznakové masce.
  - ANO ✓ ■ **sníží** počet učitelných parametrů.
  - NE ☐ **zachová** počet učitelných parametrů.
  - NE ☐ **zvýší** počet učitelných parametrů.

- Mějme batch-norm vrstvu s batch_size=20, channels=100, width=45, height=55.
  - ☐ Velikost výstupu je 20
  - ☐ Velikost výstupu je $20 \times 1 \times 45 \times 55$
  - ✓ ■ Velikost výstupu je $20 \times 100 \times 45 \times 55$
  - ☐ Velikost výstupu je 20×100
  - × ■ Velikost parametru $\gamma = 20$
  - ☐ Velikost parametru $\gamma = 100$
  - ☐ Velikost parametru $\gamma = 45 \times 55$
  - ☐ Velikost parametru $\gamma = 20 \times 100 \times 45 \times 55$