

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

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

$$\alpha_{\text{converge}} \in (0; 1/5)$$

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

$$\alpha_{\text{oscillate}} \in 1/5$$

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

$$\alpha_{\text{diverge}} \in (1/5; \infty)$$

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

$$\alpha^* = 1/10$$

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

$$N = 1$$

$$\text{rate}(\alpha^*) = 0$$

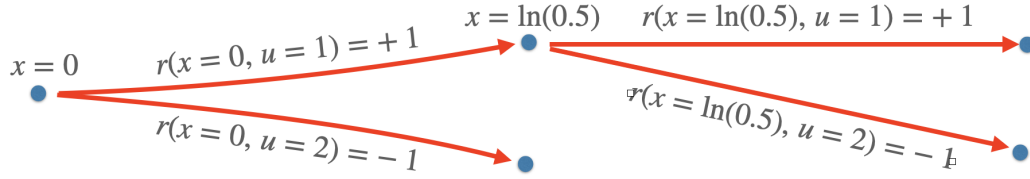
TASK 1.6 Jaká je optimální rychlost učení α^* , pro funkci $f(w_1, w_2) = 5w_1^2 + 2w_2^2 + 1$

$$\alpha^* = 1/7$$

2. Uvažujte stochastickou diskretní 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 $\mathbf{x}$ na pravděpodobnostní rozdělení dvou možných akcí $u = 1$ nebo $u = 2$.



Spočtěte následující pro $\gamma = 1$.

Tip: spočtěte si nejdříve hodnoty π_θ v jednotlivých uzlech s využitím $\sigma(-\ln(0.5)) = 2/3$:

TASK 2.1 $Q^{\pi_\theta}(x = 0, u = 1) = 4/3$

$Q^{\pi_\theta}(x = 0, u = 2) = -1$

TASK 2.2 $V^{\pi_\theta}(x = 0) = 1/6$

$V^{\pi_\theta}(x = \ln(0.5)) = 1/3$

TASK 2.3 $A^{\pi_\theta}(x = 0, u = 1) = 1$

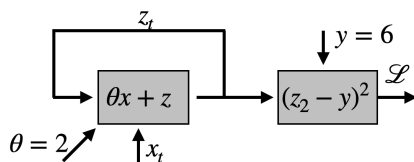
$A^{\pi_\theta}(x = 0, u = 2) = -4/3$

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).

Tip: $\sigma(x)' = \sigma(x)(1 - \sigma(x))$

$$\left. \frac{\partial \log \pi_\theta(u|x)}{\partial \theta} \right|_{\substack{x=0 \\ u=1}} \cdot A^{\pi_\theta}(x = 0, u = 1) = 0$$

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 .

Tip: Rozviňte síť v čase, abyste získali dopřednou síť a proved'te backpropagation jako obvykle.

$$\frac{\partial \mathcal{L}}{\partial \theta} = -4$$

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

4. Označte která z následujících tvrzení jsou správná (zaskrnutím čtverce):

- ☐ 'F' Zavedení skip-connection přes nějakou vrstvu zvětšuje její receptive field.
- ☐ 'F' Velikost receptive field závisí na velikosti gradientu.
- ☐ 'T' Velikost receptive field závisí na velikosti konvolučního jádra.
- ☐ 'F' Velikost receptive field závisí na batch_size (tj. velikosti batch který vstupuje do batch-norm vrstvy).
- ☐ 'F' Čím menší je receptive field tím lépe naučená síť generalizuje.
- ☐ 'T' Batch-norm omezuje explodování nebo mizení gradientu.
- ☐ 'F' Batch-norm zvyšuje pravděpodobnost saturace výstupních hodnot po sigmoid vrstvě.
- ☐ 'T' Skip-connection zefektivňuje u segmentační sítě typu U-net rekonstrukci tvarů jednotlivých objektů.
- ☐ 'F' 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 (uvažujte default hodnoty stride=1, padding=0, dilation=1):
 - ☐ 'F' **sníží** receptive field ve výstupní příznakové masce.
 - ☐ 'T' **zachová** receptive field ve výstupní příznakové masce.
 - ☐ 'F' **zvýší** receptive field ve výstupní příznakové masce.
 - ☐ 'T' **sníží** počet učitelných parametrů.
 - ☐ 'F' **zachová** počet učitelných parametrů.
 - ☐ 'F' **zvýší** počet učitelných parametrů.
- Mějme batch-norm vrstvu s batch_size=20, channels=100, width=45, height=55.
 - ☐ 'F' Velikost výstupu je 20
 - ☐ 'F' Velikost výstupu je $20 \times 1 \times 45 \times 55$
 - ☐ 'T' Velikost výstupu je $20 \times 100 \times 45 \times 55$
 - ☐ 'F' Velikost výstupu je 20×100
 - ☐ 'F' Velikost parametru $\gamma = 20$
 - ☐ 'T' Velikost parametru $\gamma = 100$
 - ☐ 'F' Velikost parametru $\gamma = 45 \times 55$
 - ☐ 'F' Velikost parametru $\gamma = 20 \times 100 \times 45 \times 55$

.

- **Feedback:** Tuto stránku odtrhněte a při odevzdání dejte na zvláštní hromádku.

Co se vám nelíbilo:

- Které **přednášky** bychom měli **odstranit/zkrátit/prodloužit/změnit**?
- Která **cvičení** bychom měli **odstranit/zkrátit/prodloužit/změnit**?
- Který **domácí úkol** máme **odstranit/zkrátit/prodloužit/změnit**?
- Něco **jiného** co bychom měli **změnit**?

Co se vám líbilo:

- Které **přednášky** bychom měli **zachovat**?
- Která **cvičení** bychom měli **zachovat**?
- Který **domácí úkol** bychom měli **zachovat**?
- Něco **jiného** co bychom měli **zachovat**?

V případě, že máte ještě dost času, nakreslete nám vtipný vánoční obrázek ;-)