VIR 2019 **Name:** ____________________

Exam test

Variant: A Points ____________________

---

1. We denote Leaky Rectified Linear Unit function with parameter α as $\mathbf{y} = \mathbf{lrelu}(\mathbf{z}, \alpha)$. The function maps single input $\mathbf{z}$ on single output value $\mathbf{y}$, The parameter α corresponds to its slope for negative inputs. Derive gradient

$$\frac{\partial \mathbf{lrelu}(-\mathbf{x}^\top \mathbf{x} - 1, \alpha)}{\partial \mathbf{x}} = ?$$

in any point $\mathbf{x} \in \mathtt{R}^n$.

   - Define a Leaky Rectified Linear Unit $\mathbf{lrelu}(\mathbf{x})$ activation function in pseudocode, with $\alpha = 0.1$. The function has a single argument $\mathbf{x}$ and output $\mathbf{y} = \mathbf{lrelu}(\mathbf{x})$.

   - Define the gradient of the $\mathbf{lrelu}(\mathbf{x})$ activation function in pseudocode. The function has a single argument $\mathbf{x}$ and outputs $\frac{\partial lrelu(x)}{\partial \mathbf{x}}$. Hint: Break up the function into two separate cases (if-else).

2. You are given batch of three one-dimensional training examples $x_1 = 5, x_2 = 2, x_3 = 1$.

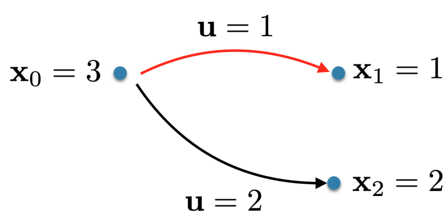
   - Compute output of the batch-norm layer with learnable parameters $\gamma = 6$, $\beta = -1$.

- Compute gradient of the batch-norm layer with respect to the parameter β.
  Hint: output of the batch-norm layer for this batch is three-dimensional.

3. Consider MDP consisting of three states $\mathbf{x}_0 = 3, \mathbf{x}_1 = 1, \mathbf{x}_2 = 2$ and two types of actions $\mathbf{u} = 1$ and $\mathbf{u} = 2$, see image below. Agent selects action $\mathbf{u}$ in the state $\mathbf{x}$ according the following stochastic policy

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

with scalar parameter $\theta = 2$. This policy maps one-dimensional state $\mathbf{x}$ on the probability distribution of two possible actions $\mathbf{u} = 1$ or $\mathbf{u} = 2$.



Consider trajectory-reward function defined as follows:

$$r(\tau) = \sum_{\mathbf{x}_i \in \tau} \frac{1}{\mathbf{x}_i}$$

Given training trajectory $\tau = [(\mathbf{x}_0 = 3), (\mathbf{u} = 1), (\mathbf{x}_1 = 1)]$, which consists of the single transition (outlined by red color), estimate:

- Policy gradient

$$\frac{\partial \log \pi_\theta(\mathbf{u}|\mathbf{x})}{\partial \theta}\Big|_{\substack{\mathbf{x} = \mathbf{x}_0 \\ \mathbf{u} = \mathbf{u}_0}} \cdot r(\tau) =$$

- Updated weights with learning rate $\alpha = 1$

$\theta^{\text{updated}} =$