2.2 a)

$H_m = WH_{m-1} + UX_m$

$= W(WH_{m-2} + UX_{m-1}) + UX_m$

$= W^2H_{m-2} + WUX_{m-1} + UX_m$

$= W^3H_{m-3} + W^2UX_{m-2} + WUX_{m-1} + UX_m$

$= \dots$

since $H_0 = 0$:

$$\boxed{H_m = \sum_{t=1}^{m} W^{m-t} U X_t}$$

this means

$$Y_m = V \cdot \sum_{t=1}^{m} W^{m-t} U X_t$$

$$\left(\boxed{Y_m = \sum_{t=1}^{m} V \cdot W^{m-t} U X_t} \right)_K \rightarrow \begin{cases} 1 \le m \le n \\ 1 \le K \le b \end{cases}$$

$$\equiv$$

$$g(\{X_t\}_{t=1}^{n}, W, U, V)_{Km}$$

2.2b)

$\rightarrow g(\langle X_t \rangle_{t=1}^{n}, W, U, V)_{K,m} = \left(\sum^{m} V W^{m-t} U X_t\right)_K$

$\rightarrow f(z)_i = \sum_{j \in \mathcal{J}} \theta_{ij} z_j, \quad \forall z \in \mathbb{R}^{\mathcal{J}}, \forall i \in \mathcal{Z}$

$X \equiv z \Rightarrow$ Given that $x_t \in \mathbb{R}^d \rightsquigarrow X \in \mathbb{R}^{d \times n}$

this implies:

$$\boxed{z_{(u,v)} = x_{u,v}, \quad \mathcal{J} = \{\langle u,v \rangle \mid 1 \leq u \leq d, 1 \leq v \leq n\}}$$

$g(\langle x_t \rangle_{t=1}^{n}, W, U, V) \equiv f(z) \Rightarrow$

$$\boxed{\mathcal{I} = \{(K,m) \mid 1 \leq K \leq b, 1 \leq m \leq n\}}$$

so: $g(\langle x_t \rangle_{t=1}^{n}, W, U, V)_{Km} = \sum_{t=1}^{m} (V W^{m-t} U X_t)_K = (Y_m)_K$

$$= \sum_{t=1}^{m} (V W^{m-t} U X_t)_K \quad \begin{cases} 1 \leq K \leq b \\ 1 \leq m \leq n \end{cases}$$

$$= \sum_{v=1}^{m} V W^{m-v} U X_{*v}$$

$$= \sum_{v=1}^{m} V W^{m-v} \left(\sum_{w=1}^{d} U_{jw} \cdot X_w\right)_j \quad \forall j = 1, \dots, d$$

$$= \underbrace{\sum_{v=1}^{m} \sum_{w=1}^{d} V W^{m-v} U_{jw}}_{\sum_{j \in \mathcal{J}} \theta_{ij}} \underbrace{X_{uv}}_{z_j} \quad \forall j = 1, \dots, d$$

2.3

a) GCN $\begin{cases} \hat{B}=0, (i,i)\in\mathcal{E}, \forall i\in V \\ W_{ij}\in\{0 | \exists j,i)\in\mathcal{E}\}^{-1} \end{cases}$

K diff topologies which propagate messages (relational GCN):

$$V' \leftarrow \sigma\left(\sum_{k=1}^{K} W^k V B^k\right)$$

→ Let W^k be the matrix representation for all Kernel values defined in the CNN $\in \mathbb{R}^{(K\times3)\times(h_y\times h_x)}$ } 2 dimensional equiv

→ Let V be the vector equivalent ~~the~~ to the ~~=~~ input variables X with ~~#~~ ~~$\forall u$~~ $\forall u \in \mathbb{R}^{f}$ $f=(H-K_y+1)\times(W_x-K+1)$

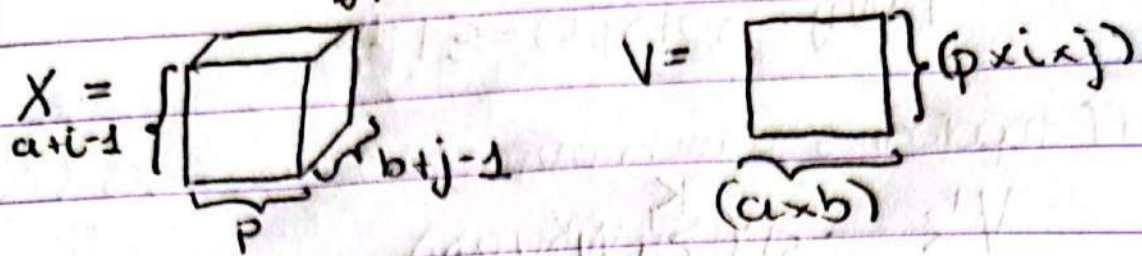
1 dimensional equivalent to input X

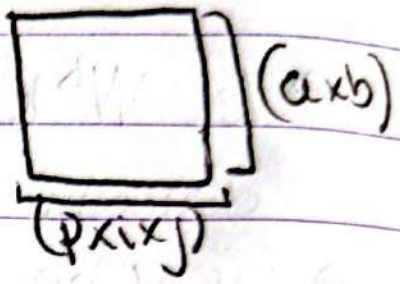
→ Let B^k be the ~~queen~~ params to optimize with the adewate optimizer $\in \mathbb{R} \in (H-K_y+1)\times(W-K+1)^T$

1 dimensional

This GCN ~~#~~ representation would be equivalent to CNN presented

2.3 b) $V\binom{a}{b}\binom{p}{i}{}_{j} = X_{p,a+i-1,b+j-1}$



Weight matrix W on $\binom{p}{i}{}_{j}\binom{a}{b} =$ 

Defining the matrix with a single-head attention layer ~~as~~ would imply that we don't know the exact topology to use for the ~~problem~~ suggested.

Given that:

$$V' \longleftarrow \sigma\left(\sum_{K=1}^{K} W^K V B^K\right)$$

if $K=1$

$$V' \longleftarrow \sigma(WVB)$$

So

$$\boxed{W \Leftarrow \sigma\left(B^{-1} V^{-1} V'\right)}$$

2.4 a) We know that: $V^*(s_0) = r(s_0) + \gamma[V(s')\ s_0]$:

Best policy without memory:

Given that $p(c=\text{green}) = 1/3$ the best policy is the one that goes to s_8 [$p(c_6=\text{red}) = 2/3$) and the policy value is (+1 reward, −1 reward)

$$V^*(s_0) = 2/3 \cdot \gamma^4 - 1/3\ \gamma^4 \quad \text{(since we do 4 steps)}$$

$$\boxed{= \gamma^4/3}$$

Best policy with memory: Given that we have a memory we should take ~~other~~ advantage of state s_2 so:

$$\boxed{V^*(s_0) = \gamma^8}$$ since you first go to s_2 in order to know what is the value of c and be sure were is the positive value

The range of values for which γ is better a memory-less policy is $[0 \ldots \sqrt[4]{1/3}]$

since: $\gamma^4/3 = \gamma^8 \rightarrow$ equality

$$1 = 3\gamma^4$$

$$\boxed{\gamma = \sqrt[4]{1/3}}$$

2.4 b) The smallest set of past observations would be a set of 4, since the decision ~~white~~ in which the policy is going to influence s_4 is related to state s_4 and from s_2 to s_4 there are 4 steps. When you are in ~~all other~~ states s_8, s_7 the policy is clear always. ($s_5 \rightarrow s_6$ & $s_7 \rightarrow s_8$)

2.4 c)

Without memory

There is no effect in the policy since the action associated into a state always remains the same. Analyze both cases

50% go in desired direction → the policy remains the same

50% chance you stay in current state → you are not in the desired ~~stay~~ state but the action to go there remains the same

With memory

The effect on the policy is the same, it only affects the capacity of the memory that should increase in order to provide V^*.

2.5 a)

$$\nabla_\theta E[f(s,a) \mid a \sim \pi_\theta(a|s)] = \nabla_\theta \int f(s,a)\, p(s,a)\, da$$

$$= \int f(s,a)\, \nabla_\theta p(s,a)\, da$$

since $\nabla_\theta \ln(p,sa) = \frac{1}{p(s,a)} \nabla_\theta p(s,a)$

$$= \int f(s,a)\, \nabla_\theta \ln p(s,a)\, p(s,a)\, da$$

$$= E[f(s,a)\, \nabla_\theta \ln \pi_\theta(s|a) \mid a \sim \pi_\theta(s,a)]$$

2.6 b) $\nabla_\theta E_{\pi_\theta}[R_t \mid s_t a_t] = \nabla_\theta E_{\pi_\theta}[r(s_t,a_t) + \gamma R_{t+1} \mid s_t a_t]$

$$= \nabla_\theta (E_{\pi_\theta}[r(s_t,a_t \mid s_t a_t] + \gamma E[R_{t+1} \mid s_t a_t]$$

$$= \nabla_\theta E_{\pi_\theta}[r(s_t,a_t) \mid s_t a_t] + \nabla_\theta \gamma E[R_{t+1} \mid s_t a_t]$$

$$= \nabla_\theta E_{\pi_\theta}[r(s_t,a_t) \mid s_t a_t] + \gamma E[R_{t+1} \nabla_\theta \ln \pi_\theta(a_t|a_t) \mid s_t a_t]$$

$$\nabla_\theta E_{\pi_\theta}[r(s_t,a_t) \mid s_t a_t] = \nabla_\theta \int r(s_t,a_t)\, p(s_t,a_t)\, da$$

$$= \int r(s_t,a_t)\, \nabla_\theta p(s_t,a_t)\, da$$

$$= \int r(s_t,a_t)\, \nabla_\theta p(s_t,a_t)\, p(s_t,a_t)\, da$$

$$= E_{\pi_\theta}[r(s_t,a_t)\, \nabla_\theta \ln \pi_\theta(a_t|s_t) \mid s_t a_t]$$

$$= E_{\pi_\theta}[r(s_t,a_t)\, \nabla_\theta \ln \pi_\theta(a_t|s_t) \mid s_t a_t] +$$

$$\gamma E_{\pi_\theta}[R_{t+1} \nabla_\theta \ln \pi_\theta(a_t|s_t) \mid s_t a_t]$$

2.5 c) $\nabla_\theta J[\pi_\theta] = \nabla_\theta E_{\pi_\theta}[R_0]$

~~$= \nabla_\theta E_{\pi_\theta}[\sum_{k=1}^{n-1} r^{k-t} r(s_0, a_0)]$~~

~~$= \nabla_\theta E[\sum_{k=1}^{n-1} r^{k-t} r(s_0|a_0) \pi_\theta(a_0|s_0)]$~~

~~$\Rightarrow \nabla_\theta \sum_{k=1}^{n-1} E[r^{k-t} r(s_0|a_0) \pi_\theta(a_0|s_0)]$~~

$= \nabla_\theta E_{\pi_\theta}[R_0 + rR_1 + r^2R_2 + \dots$

$= E_{\pi_\theta}[\nabla_\theta R_0 + \nabla_\theta rR_1 + \nabla_\theta r^2 R_2 \dots]$

$\Rightarrow$ this can be also expesse as

$= E_{\pi_\theta} \nabla_\theta E_{\pi_\theta}[\sum_{t=0}^{n-1} r^t R_t]$

$= E_{\pi_\theta}[\sum_{t=0}^{n-1} \nabla_\theta r^t R_t]$ ↗ Given previously exercise 2.5 a)

$= E_{\pi_\theta}[\sum_{t=0}^{n-1} r^t R_t \nabla_\theta \ln \pi_\theta(a_t|s_t)]$