COMP4211 - Machine Learning

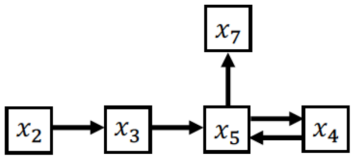
Fall 2024, HKUST

Problem Set

Name: Shawn Darren S. Chua **Student ID:** 20946705

Problem 1.

(1)



Write your answer ⇒

Incoming Matrix: A^I

	2	3	5	4	7
2	0	0	0	0	0
3	1	0	0	0	0
5	0	1/2	0	1/2	0
4	0	0	1	0	0
7	0	0	1	0	0

Outgoing Matrix: A^O

	2	3	5	4	7
2	0	1	0	0	0
3	0	0	1	0	0
5	0	0	0	1/2	1/2
4	0	0	1	0	0
7	0	0	0	0	0

(2) $h^{l-1^\top} \in \mathbb{R}^d$ and $h^{l-1^\top}\mathbf{W}^I \in \mathbb{R}^d$, so $\mathbf{W}^I \in \mathbb{R}^{d\times d}$

$h^{l-1^\top} \in \mathbb{R}^d$ and $h^{l-1^\top}\mathbf{W}^O \in \mathbb{R}^d$, so $\mathbf{W}^O \in \mathbb{R}^{d\times d}$

Thus $\mathbf{W}^I, \mathbf{W}^O$ both have d^2 parameters.

(3) $a_i^l \in \mathbb{R}^{2d\times 1}$ and $\mathbf{W}_z a_i^l \in \mathbb{R}^{d\times 1}$, so $\mathbf{W}_z \in \mathbb{R}^{d\times 2d}$

$a_i^l \in \mathbb{R}^{2d\times 1}$ and $\mathbf{W}_r a_i^l \in \mathbb{R}^{d\times 1}$, so $\mathbf{W}_r \in \mathbb{R}^{d\times 2d}$

$a_i^l \in \mathbb{R}^{2d\times 1}$ and $\mathbf{W}_h a_i^l \in \mathbb{R}^{d\times 1}$, so $\mathbf{W}_h \in \mathbb{R}^{d\times 2d}$

Thus $\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_h$ all have $2d^2$ parameters.

$h_i^{l-1} \in \mathbb{R}^{d\times 1}$ and $\mathbf{U}_z h_i^{l-1} \in \mathbb{R}^{d\times 1}$, so $\mathbf{W}_z \in \mathbb{R}^{d\times d}$

$h_i^{l-1} \in \mathbb{R}^{d\times 1}$ and $\mathbf{U}_r h_i^{l-1} \in \mathbb{R}^{d\times 1}$, so $\mathbf{W}_r \in \mathbb{R}^{d\times d}$

$r_i^l \odot h_i^{l-1} \in \mathbb{R}^{d\times 1}$ and $\mathbf{U}_h \left(r_i^l \odot h_i^{l-1}\right) \in \mathbb{R}^{d\times 1}$, so $\mathbf{W}_h \in \mathbb{R}^{d\times d}$

Thus $\mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_h$ all have d^2 parameters.

(4)

(5) $\mathbf{W}^I, \mathbf{W}^O$ have a total of $2d^2$ parameters.
$\mathbf{b}^I, \mathbf{b}^O \in \mathbb{R}^d$ have a total of $2d$ parameters.
$\mathbf{W}_z, \mathbf{W}_r, \mathbf{W}_h$ have a total of $6d^2$ parameters.
$\mathbf{U}_z, \mathbf{U}_r, \mathbf{U}_h$ have a total of $3d^2$ parameters.
$\mathbf{W}_{q1}, \mathbf{W}_{q2}$ have the same shapes as $\mathbf{W}_{k1}, \mathbf{W}_{k2}$, so they have a total of $6d^2$ parameters.
Summing these up, we get $2d^2 + 2d + 6d^2 + 3d^2 + 6d^2 = 17d^2 + 2d = 170200$ parameters.

Problem 2.

(a)

(b)