Problem 1. (Reinforcement learning) . . . /3 **points**

You want to use reinforcement learning to control a drone. You model the drone as a Markov decision process

1. Each of the drone's 4 motors are controlled by a voltage $V \in [0, 12]$. What is the action space of the Markov decision problem? **(1 point)**

   $A = \dots\dots\dots$
   Solution: $A = [0, 12]^4$. $A = \mathbb{R}^4$ is worth 0.5 point

2. Let's $s \in \mathbb{R}^{12}$ denote the state of the drone, and consider a linear policy $\pi : \mathbb{R}^{12} \to A$ such that $\pi(s) = Ks$, where K is a matrix. Write down the dimensions of K. **(1 point)**

   $K \in \dots\dots\dots$
   Solution: $K \in \mathbb{R}^{4\times 12}$

3. Consider the following trajectories where $s_t^i \in \mathbb{R}^{12}$ and $a_t^i \in A$, for $i \in \{1, 2, 3\}$:

trajectory 1	s_0	a_0^0	s_1^0	a_1^0	s_2^0	a_2^0	s_3^0	a_3^0	s_T^0
trajectory 2	s_0	a_0^1	s_1^1	a_1^1	s_2^1	a_2^1	s_3^1	a_3^1	s_T^1
trajectory 3	s_0	a_0^2	s_1^2	a_1^2	s_2^2	a_2^2	s_3^2	a_3^2	s_T^2

   You want to train your agent to minimize the function $J(s_0, \pi) = \frac{1}{2}(s_T - s_G)^2$ where s_G is a goal state and T is a target time. Write down the empirical estimate of $J(s_0, \pi)$ based on the 3 trajectories above: **(1 point)**

   $J(s_0, \pi) \approx \dots\dots\dots\dots\dots\dots$
   Solution: $J(s_0, \pi) \approx \frac{(s_T^0 - s_G)^2 + (s_T^1 - s_G)^2 + (s_T^2 - s_G)^2}{6}$

Problem 2. (Convolutional Neural Network) . . . . . . /1.5 **points**

Fill in the empty blocks in the convolution of X with a 2×2 filter w and a bias $b = 2$. **(1.5 points)**

0	1	1
0	1	0
1	1	1

X

1	0
0	1

w
Bias: $b = 2$

3	3
3	4

$w * X$

Problem 3. (Recurrent Neural Network) ....../2.5 **points** Suppose that we have a data set containing temperature, humidity, and wind measurements over a period of T days in Lausanne in Lausanne and we want to develop a recurrent neural network to forecast a scalar temperature y_{T+1} value for day $T+1$. The recurrent neural network has the following structure (tanh is applied coordinate wise)

$$s_{t+1} = \tanh\left(W_{ss}s_t + W_{sa}a_t\right)$$
$$\hat{y}_t = W_{os}s_t,$$

where the hidden state has dimension 30. There are no bias terms.

1. What are the dimensions of s_t, a_t, y_t, W_{ss}, W_{sa}, and W_{os}? **(1.5 points)**

   $s_t \in \ldots\ldots\ldots\ldots\ldots,\quad a_t \in \ldots\ldots\ldots\ldots\ldots,\quad \hat{y}_t \in \ldots\ldots\ldots\ldots\ldots$

   $W_{ss} \in \ldots\ldots\ldots\ldots\ldots,\quad W_{sa} \in \ldots\ldots\ldots\ldots\ldots,\quad W_{os} \in \ldots\ldots\ldots\ldots\ldots$

   Solution: The dimensions are: $s_t \in \mathbb{R}^{30}$, $a_t \in \mathbb{R}^3$, $\hat{y}_t \in \mathbb{R}$, $W_{ss} \in \mathbb{R}^{30\times30}$, $W_{sa} \in \mathbb{R}^{30\times3}$, and $W_{os} \in \mathbb{R}^{1\times30}$

2. What terms are the trainable parameters? **(0.5 point)** ....

   Solution: The trainable parameters are W_{ss}, W_{sa}, and W_{os}.

3. Write the mean-squared loss: **(0.5 point)** $L =$ ....
   Solution:

$$L = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

Problem 4. (Naive Bayes) ....../3 **points**

You receive a lot of spam emails. You decide to program your own spam filter based on word occurences using the Naive Bayes classifier. Assume that you collected the following regular and spam mails to train the classifier, and only two phrases are informative for this classification: "send money" (x_1) and "EPFL" (x_2).

Data	Spam (y)	"send money" (x_1)	"EPFL" (x_2)
email 1	1	1	0
email 2	1	1	0
email 3	1	0	1
email 4	0	0	1
email 5	1	1	0
email 6	0	1	1
email 7	0	0	0
email 8	1	1	0

1. Given the above dataset, what is the probability that an email chosen at random is spam? What is the probability (denoted as P) that the email contained "send money" given that it was spam? What is the probability that the email contained "EPFL" given that it was not spam? (**1.5 points**)

$P(y = 1) = \ldots\ldots\ldots$ $P(x_1 = 1|y = 1) = \ldots\ldots\ldots$ $P(x_2 = 1|y = 0) = \ldots\ldots\ldots\ldots$ *Solution:*

$P(y = 1) = \frac{5}{8} = 0.625$
$P(x_1 = 1|y = 1) = \frac{4}{5} = 0.8$
$P(x_2 = 1|y = 0) = \frac{2}{3} \approx 0.67$

2. You receive a new email which contains both "send money" and "EPFL" and want to compute the probability that it is not a spam email $y = 0$. Fill in the blanks in the formula below (**0.5 points**)

$$P(y = 0|x_1 = 1, x_2 = 1) = \frac{P(x_1 = \quad |y = \quad)P(x_2 = \quad |y = \quad)P(y = \quad)}{P(x_1 = 1, x_2 = 1)}$$

Solution: $P(y = 0|x_1 = 1, x_2 = 1) = \frac{P(x_1=1|y=0)P(x_2=1|y=0)P(y=0)}{P(x_1=1,x_2=1)}$

3. Using the previous result and the gathered data, compute the probability for the email containing both "send money" and "EPFL" to be spam. (**1 points**)

$$P(y = 1|x_1 = 1, x_2 = 1) = \frac{\ldots\ldots\ldots\ldots\ldots\ldots}{P(x_1 = 1, x_2 = 1)}$$

Solution:

$$\begin{aligned} & P(y = 1|x_1 = 1, x_2 = 1) \\ &= \frac{P(x_1 = 1|y = 1)P(x_2 = 1|y = 1)P(y = 1)}{P(x_1 = 1, x_2 = 1)} \\ &= \frac{\frac{4}{5}\frac{1}{5}\frac{5}{8}}{P(x_1 = 1, x_2 = 1)} \\ &= \frac{\frac{1}{10}}{P(x_1 = 1, x_2 = 1)} \end{aligned}$$

The answere $P(y = 0|x_1 = 1, x_2 = 1) = 1/(12P(x_1 = 1, x_2 = 1))$ is considered to be worth 1 point as well due to the typo.