ML EXAMS MAY 12/04/2018

EX. 1

1. given an dataset D and an hypothesis space H a model is overfitting when exist another hypothesis $h' \in H$ such that:

$error_D(h) > error_D(h')$ and $error_S(h) < error(h')$

two nodes one

2. If we have two decision trees and one is deeper respect to other for sure we have overfitting we have two possible methods to avoid

I. REDUCE error pruning: using validation dataset, we prune the node that doesn't improve validation accuracy

II. rule post prune: if the accuracy after pruning is greater respect before we have overfitting

EX 2

1. naive bayes classifier uses conditional independence to approximate the solutions:

x is conditionally independent from Y given Z

$$P(X, Y | Z) = P(X | Y, Z) P(Y | Z) = P(X | Z) P(Y | Z)$$

assume target function $f: X \to V$ where each instances of x are describe by $\langle a_1 \dots, a_n \rangle$

$$\underset{v_j \in V}{\operatorname{argmax}} P(v_j | x, D) = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | a_1 \dots, a_n, D) = V_{MAP}$$

with bayes assumption

$$P(a_1 \dots, a_n | v_j, D) = \prod_j P(a_i | v_j, D)$$

$$V_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j | D) \prod_i P(a_i | v_j | D)$$

2. in a classification task: the target function $f: X \to Y$ with $X = \{$ title x author x abs x site $\}$ and $Y = \{ML, KR, PL\}$ the dataset is $D = \{(d_i, y_i)\}_{i=1}^{N}$

$$V_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v | D) \prod_i P(d_i | V, D)$$

we know $P(d_i | V, D) = \prod_{i=2}^{L} P(d_i = w_i | v, D)$ where $L = length(d_i)$ and $P(d_i = w_k | V, D)$ probability that the word occours in the document i

$$\hat{P}(d_i | V, D) = TF_{ij+2} / TF_{i+1}$$

Ex. 3

1. because we have two classes {0, 1} consider a logistic regression that is a probabilistic discriminative model based on maximum likelihood

   likelihood

   likelihood

$$P(t|\vec{w}) = \prod_{n=1}^{N} y_n^{t_n} (1-y_n)^{(1-t_n)} \quad \text{with } y_n = P(C_1|\vec{x}_n)$$

2. the parameters that the model has to learn are a_i and a_i

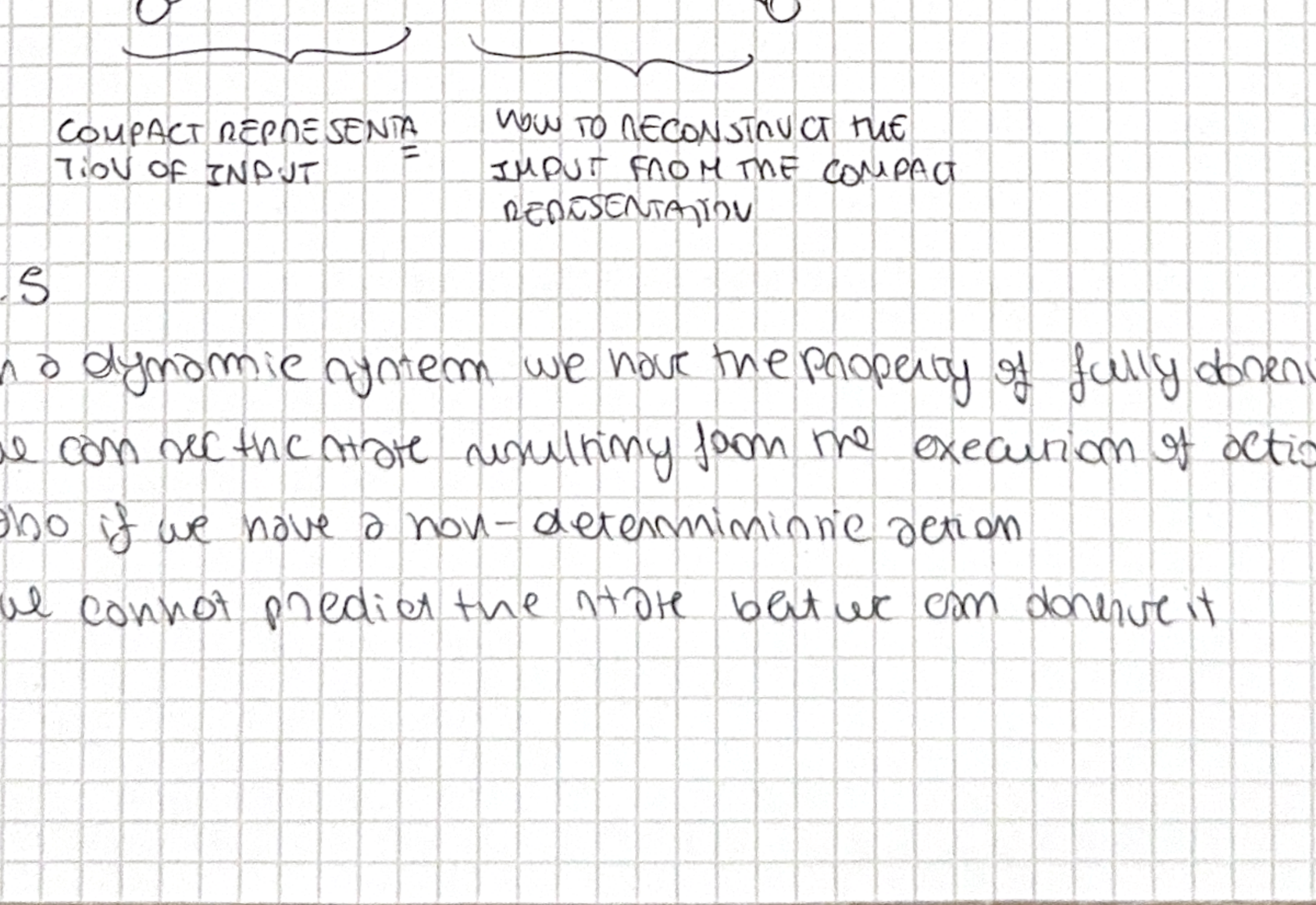
3. a suitable error function will from entropy loss

$$E(\vec{w}) = -\ln(P(t|\vec{w}) = -\sum_{n=1}^{N} [t_n \ln(y_n + (1-t_n)\ln(1-y_n)]$$

Ex. 4

1. an autoencoders is combination of two ANN a decoder and encoder the training is based on reconstruction loss.
   An auto encoders take in input and output the same sample x_n from the dataset (is used to deal with data far away from Gaussian distribution)

2.



Ex. 5

1. in a dynamic system we have the property of fully observability if we can see the state resulting from the execution of action
   also if we have a non-deterministic action we cannot predict the state but we can observe it

2) in MDP we have $\langle X, A, \delta, r \rangle$

X: finite set of state A: finite set of actions

$\delta: X \times A \to X$ is the transition function and $X \times A \to \mathbb{R}$: r → reward function

in POMDP we have $\langle X, A, Z, \delta, r, O \rangle$ where

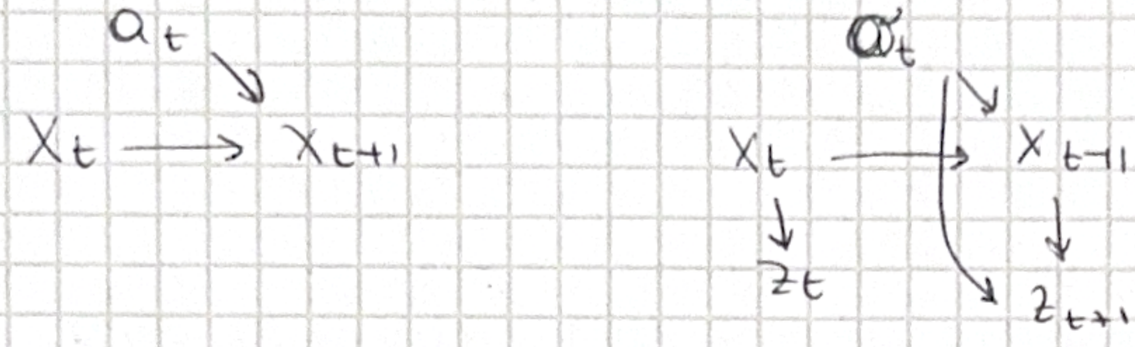
X: finite set of states A: finite set of actions Z: finite set of

$\delta: X \times A \to X$ transition function $r: X \times A \to \mathbb{R}$ is the observation

reward function and O is the probability of observation and

$P(x_0)$ is the probability of the initial state

Ex 3.



EX. 6

1. $\dim(W_1) = 128 \cdot 10 = 1280$ $\dim(W_2) = 50 \cdot 10 = 500$

2 $y(x) = w^T \max(0, w^T x + c) + b$ $h = g(w^T x + c)$

$g(a) = \max(0, a) \to ReLU$

ML EXAMS 18/01/2018

EX. 1

input image = $1242 \times 378 \times 3$ (RGB) kernel = 5×5 pad = 2 stride = 1

$W_{out} = \frac{W_{in} - W_k}{s} + 2P + 1 = 1242$

$d_{out} = \frac{d_{in} - W_{kernel}}{s} + 2P + 1 = 378$

~~number of trainable parameters: $5 \times 5 \times 3 \times 64$~~

$W_{out} = \frac{W_{in} - d_{pool}}{2} + 1 = 621$

$d_{out} = \frac{378 - d_{pool}}{2} + 1 = 189$

~~$621 \times 189 \times 64$~~

~~$310 \times 94 \times 128$~~

$W_{out} = 310$ $h_{out} = 84$ $C_1 \to 1242 \times 378 \times 64$

$W_{out} = 78$ $H_{out} = 24$ $P_1 \to 621 \times 189 \times 64$

$C_2 \to 310 \times 84 \times 128$ $P_2 \to 78 \times 24 \times 128$

2) 1 ✗ params = $3 \times 5 \times 5 + 64 = 1164$

2 ✗ params = $64 \times 3 \times 3 * 128 + 128 = 73856$