EXAM 23.06.20

Question 1

1). The output is the number of occurrences of "a" squared. This is a regression task. The target function is $f: X \to Y$ with $X = \{A_1 \times A_2 \times \dots A_N\}$ and $Y = \mathbb{R}$.
The model can be simply written as: $\phi(x) = \sum_{n=1}^{N} \mathbb{I}(x_n = a)^2 \Rightarrow y(x) = W^T \phi(x)$

2). We can introduce a more complex representation: we use a k-hot vector $V \in Z_{26}^{+}$ with the correlation between a number and a letter of the alphabet.
$y = \sum_{n=1}^{N} (W_n^T V_n) V_n = (W^T V^T) V$
If we use the kernel function $\phi(x, x') = x^T x' \Rightarrow y = W^T \phi(V)$
3). $y = W_n^T k(V^T, V) = \sum_{n=1}^{N} \alpha_n k(x_n, x)$ with $k(x, y) = x \cdot y$

Question 2

1). $S = [0|011]$ $S = [0|101]$ $S = [1|110]$ $S = [1|011]$
$S = [0|110]$ $S = [1|111]$ $S = [1|101]$
$P(S) = \frac{1}{7}$

2). $P(S|b_0 = 0) = \frac{1}{7} \cdot \frac{3}{7} = \frac{3}{49}$ $S_{MAP}^1 = [0|011]$ $S_{MAP}^2 = [0|110]$ $S_{MAP}^3 = [0|101]$

3). $S_{ML}^1 = [1|111]$ $S_{ML}^2 = [1|110]$ $S_{ML}^3 = [1|101]$
$P(b_2 = 1 | b_0 = b_1 = 1) = \frac{2}{3}$

Question 3

1). $P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x)} = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)} = \frac{1}{1 - e^{-\alpha}} = \sigma(\alpha)$
with $\alpha = \ln \frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)} = \ln \frac{N(x|\mu_1, \Sigma_1)P(C_1)}{N(x|\mu_2, \Sigma_2)P(C_2)} = W_0 + W^T x$
$W_0 = -\frac{1}{2}\mu_1^T \Sigma_1^{-1} \mu_1 + \frac{1}{2}\mu_2^T \Sigma_2^{-1} \mu_2 + \ln \frac{P(C_1)}{P(C_2)}$ $W = \Sigma^{-1}(\mu_1 - \mu_2)$
$P(C_1|x) = \sigma(W_0 + W^T x)$
2). $\mu_i = \begin{pmatrix} \dots \\ \dots \end{pmatrix}$ so the number of unknown parameters is 2.

$\Sigma_i = \begin{pmatrix} \dots & \dots \\ \dots & \dots \end{pmatrix}$ So the number of unknown parameters is 3.

Size $= (2 \cdot 2) + (3 \cdot 3) + 1 = 11$

Question 4

We have a single state MDP $= \langle \{x_0\}, A, \delta, r \rangle$ with $\delta: X \times A \to X$ transition functi... and $\delta(a_i) = x_0$ $\forall a_i \in A$ and $r: X \times A \to \mathbb{R}$ reward function. We want to find th... optimal policy $\pi^*(x_0)$ and we have 4 different cases:

1). If $r(a_i)$ is deterministic and known: $\pi^*(x_0) = \underset{a_i \in A}{\operatorname{argmax}}\, r(a_i)$

2). If $r(a_i)$ is deterministic and unknown:

We execute each action $a_i \in A$ and we collect the reward $r(a_i)$

$\pi^*(x_0) = a_i$ with $i = \underset{i=1 \dots |A|}{\operatorname{argmax}}\, r(i)$

3). If $r(a_i)$ is non-deterministic and known: $\pi^*(x_0) = \underset{a_i \in A}{\operatorname{argmax}}\, E[r(a_i)]$

4). If $r(a_i)$ is non-deterministic and unknown:

foreach i we initialize $C[i] \leftarrow 0$ and $Q_{(0)}[i] \leftarrow 0$

foreach time $t = 1 \dots T$:

We choose an index $\hat{i}$ such that $a(\hat{i}) = a(t)$

We execute the action and we collect the reward

We increment $C[\hat{i}]$

$Q_{(t)}[\hat{i}] \leftarrow \frac{1}{C[\hat{i}]}\left(\bar{r}(t) + (C[\hat{i}] - 1)\, Q_{(t-1)}[\hat{i}]\right)$

$\pi^*(x_0) = a_i$ with $i = \underset{i=1 \dots |A|}{\operatorname{argmax}}\, Q_{(T)}[i]$

2). $Q_n(a_i) \leftarrow Q_{n-1}(a_i) + \alpha[\bar{r} - Q_{n-1}(a_i)]$ with $\alpha = \frac{1}{1 + V_{n-1}(a_i)}$

$V_{n-1}(a_i)$ = number of executions of the action a_i up to time $n-1$.

Question 5

1). $W_{out} = \frac{W_{in} - W_k + 2P}{S} + 1 = \frac{256 - 2}{2} + 1 = 128$

The features map size is $128 \times 128 \times 64$.

2). We can have a convolutional layer with:

8×8 kernel padding $= 4$ stride $= 2$

Question 6

We can use boosting: $y_B(x) = \frac{1}{M}\sum_{k=1}^{M} y_k(x) = \frac{1}{3}(0.912 + 0.432 + 0.444) = 0.596 \to$ class 1