EXAM 16.12.19 (K-MEANS)

Question 1

1). If $C = C_1$ AND $B = b_1$ THEN NO (A)

If $C = C_1$ AND $B = b_2$ THEN YES (B)

If $C = C_2$ AND $A = a_1$ THEN YES (C)

If $C = C_2$ AND $A = a_2$ AND $B = b_1$ THEN YES (D)

If $C = C_2$ AND $A = a_2$ AND $B = b_2$ THEN NO (E)

If $C = C_2$ AND $A = a_3$ THEN NO (F)

If $C = C_3$ THEN NO (G)

2). T is consistent with S_1 because of A.

T is consistent with S_2 because of D.

T is not consistent with S_3 because of G.

T is not consistent with S_4 because of E.

Question 4

1). The maximum a posteriori hypothesis is $h_{MAP} = \arg\max_{h \in H} \frac{P(D|h)P(h)}{P(D)} =$

$= \arg\max_{h \in H} P(D|h)P(h)$

If we assume that $P(h_i) = P(h_j)$ we can simplify and we obtain the maximum likelihood hypothesis $h_{ML} = \arg\max_{h \in H} P(D|h)$

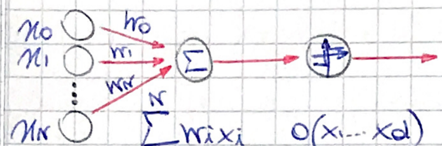
2). The Bayes Optimal Classifier is an optimal classifier, it returns always the optimal solution. Given a target function $f: X \to V$ with $V = \{V_1 \times V_2 \times \dots V_k\}$

$V_{OB} = \arg\max_{v \in V} \sum_{h_i \in H} P(v|x, D, h_i)P(h_i|D, x) = \arg\max_{v \in V} \sum_{h_i \in H} P(v|x, h_i)\, P(h_i|D)$

3). Bayes Optimal Classifier is an optimal classifier but can be used if the hypothesis space is not large or if we have analytical solutions, otherwise is not usable.

Question 3

1). The perceptron model is based on the following structure:



$$o(x_1 \dots x_d) = \begin{cases} +1 & \text{if } w_1 x_1 + \dots w_d x_d > 0 \\ -1 & \text{otherwise} \end{cases}$$

$\sum_{i=0}^{N} w_i x_i$ $\quad o(x_1 \dots x_d)$ For the moment we ignore this function.

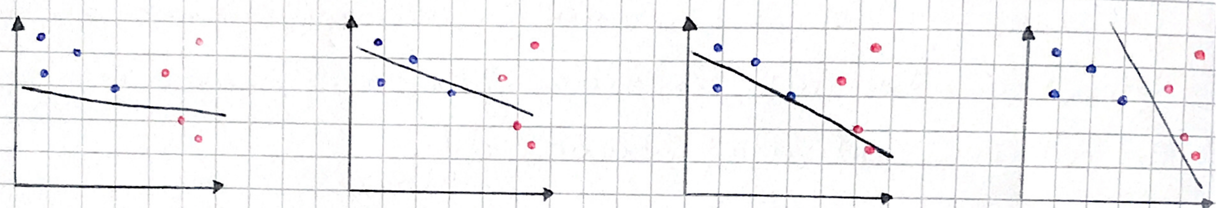
$$E(w) = \frac{1}{2} \sum_{n=1}^{N} (t_n - o_n)^2 = \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T x_n)^2$$

$\dfrac{\partial E(w)}{\partial w_i} = \sum_{n=1}^{N} (t_n - w^T x_n)(-x_{i,n})$. To find w^* we use a sequential algorithm:

$\hat{w} \leftarrow \hat{w} + \Delta w_i$ with $\Delta w_i = -\eta \sum_{n=1}^{N} (t_n - w^T x_n)(x_{i,n})$ Now we consider $o(x_1 \dots x_d)$

$$\hat{w} \leftarrow \hat{w} - \eta \sum_{n=1}^{N} (t_n - \text{sign}(w^T x_n))(x_{i,n})$$

2).



Question 4

1). $\dim(W_1) = 128 \times 10$ $\qquad \dim(W_2) = 50 \times 10 = 500$

2). $h = g(W^T x + c)$ with $g(\alpha) = \max(0, \alpha)$ ReLu function

$y(x, N) = w^T \max(0, W^T x + c) + b$

Question 6

1). An autoencoder is simply a combination of two NNs: an encoder and a decoder. The training is based on reconstruction loss and in an autoencoder we have hidden layers with reduced size (bottlenecks). Given a dataset $\{x_n\}$ an autoencoder is trained with x_n both in input and in output.

2).