ML EXAMS 13/12/2017

Question 1

1. this is a binary classification task where the target function is $f: X \to \{yes, NO\}$ with $X = \{$furniture, NR_rooms, New_Kitchen$\}$ and dataset is $D = \{(x_i, y_i)_{i=1}^{5}\}$

2. because is a binary classification problem and for solve this task we use decision tree for solve this task we can use ~~two strategies~~

ID3 that select variable the maximize INFORMATION GAIN

- INFORMATION GAIN = ENTROPY(S) - $\sum_{v \in V}$ |Sv|/S $\log_2$ (ENTROPY(Sv))

ENTROPY(S) = $\sum_{i=1}^{N}$ - $p_i \log_2 p_i$

entropy measure the impurity of our information

~~C4.5 the that maximize inform~~

3.

ENTROPY(S) = $-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.9761$

ENTROPY(FN) = $-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.916$

ENTROPY(Fy) = $-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$

IG(S, F) = $-\frac{3}{5}$ ent(FN) - $\frac{2}{5}$ ent(Fy) + ent(S) = 0.0022

ENTROPY(NR3) = $-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right) = 0.916$

ENTROPY(NR4) = 0

IG(S, NR) = en(S) - $\frac{3}{5}$ en(NR3) - $\frac{2}{5}$ en(NR4) = 0.422

ENTROPY(NKy) = 0 ENTROPY(NKN) = 0

IG(S, NK) = en(S) - $\frac{4}{5}$ ent(NKN) - $\frac{1}{5}$ ent(NKy) = 0.171

Question 2

1. given a classification problem with a dataset D and hypothesis space H we are interested $P(h|D)$.

$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$ we define Maximum A Posteriori (MAP)

hypothesis as $h_{MAP} = \underset{h \in H}{\operatorname{argmax}} \frac{P(D|h)P(h)}{P(D)}$

if we assume $P(h_j) = P(h_i)$ we can simplify and obtain:

$h_{ML} = \underset{h \in H}{\operatorname{argmax}} P(D|h)$

2. is FALSE consider this example

$h_1(x) = +$ $h_2(x) = -$ $h_3(x) = +$

$P(D|h_1) = 0.4$ $P(D|h_2) = P(D|h_3) = 0.3$

$h_{ML} = \underset{v \in V}{\operatorname{argmax}} \sum_{h \in H} P(D|h) = \operatorname{argmax}\{0.4, 0.3, 0.3\} = +$

now we introduce the concept of BOC Bayesian Optimal Classifier

$V_{OC} = \underset{v \in V}{\operatorname{argmax}} \sum_{h_i \in H} P(v|x, h_i)P(h_i|D)$

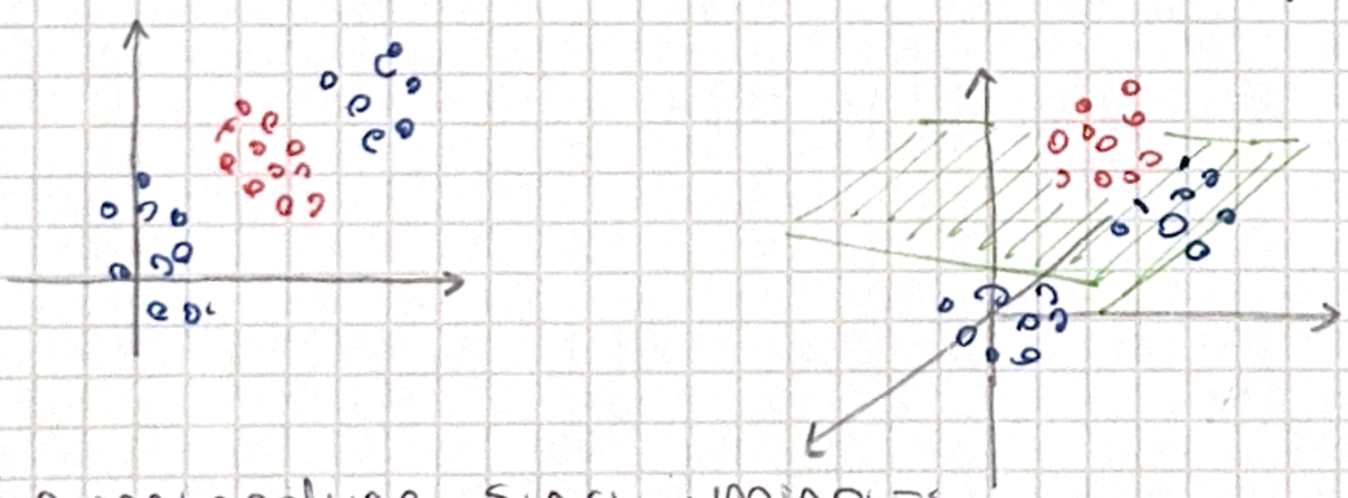
$P(+|x, h_1) = 1$ $P(+|x, h_2) = 0$ $P(+|x, h_3) = 0$

$P(-|x, h_1) = 0$ $P(-|x, h_2) = 1$ $P(-|x, h_3) = 1$

$V_{OC} = \underset{v \in V}{\operatorname{argmax}}\{(1 \cdot 0.4 + 0 + 0), (0 + 1 \cdot 0.3 + 1 \cdot 0.3)\} = -$

Question 3

1. in this case I will use a polynomial kernel function



2. we introduce SLACK VARIABLES

$\xi = 0$ if

$\xi = 0$ if points ON or INSIDE the correct margin boundary

$0 < \xi \leq 1$ if points INSIDE ~~correct margin~~ the margin but correct side

$\xi > 1$ if points on wrong side of boundary

the modified error function became $E(w) = \underset{w}{\operatorname{argmin}} \frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N} \xi_n$

and $t_n y(x_n) \geq 1 - \xi_n \quad \forall n = 1, \dots N$

Question 4

1. is a binary classification problem; function activation ~~BINARY CROSSENTROPY~~ sigmoid activation V

$$y = \delta(w^T h + b)$$

2. output unite saturates only when it gives the corect answer

Question 5

1. $w_{out} = w_{in} - \frac{w_k}{s} + 2p + 1$

$h_{out} = h_{in} - \frac{h_k + 2p}{s} + 1$

when s = stride p = padding $w_k = 3$ $d_k = 16$ $h_k = 3$

we want $w_{in} = w_{out}$ and $h_{in} = h_{out}$

$$w_{in} = -\frac{w_k + 2p + s}{s-1} \qquad h_{in} = -\frac{h_k + 2p + s}{s-1}$$

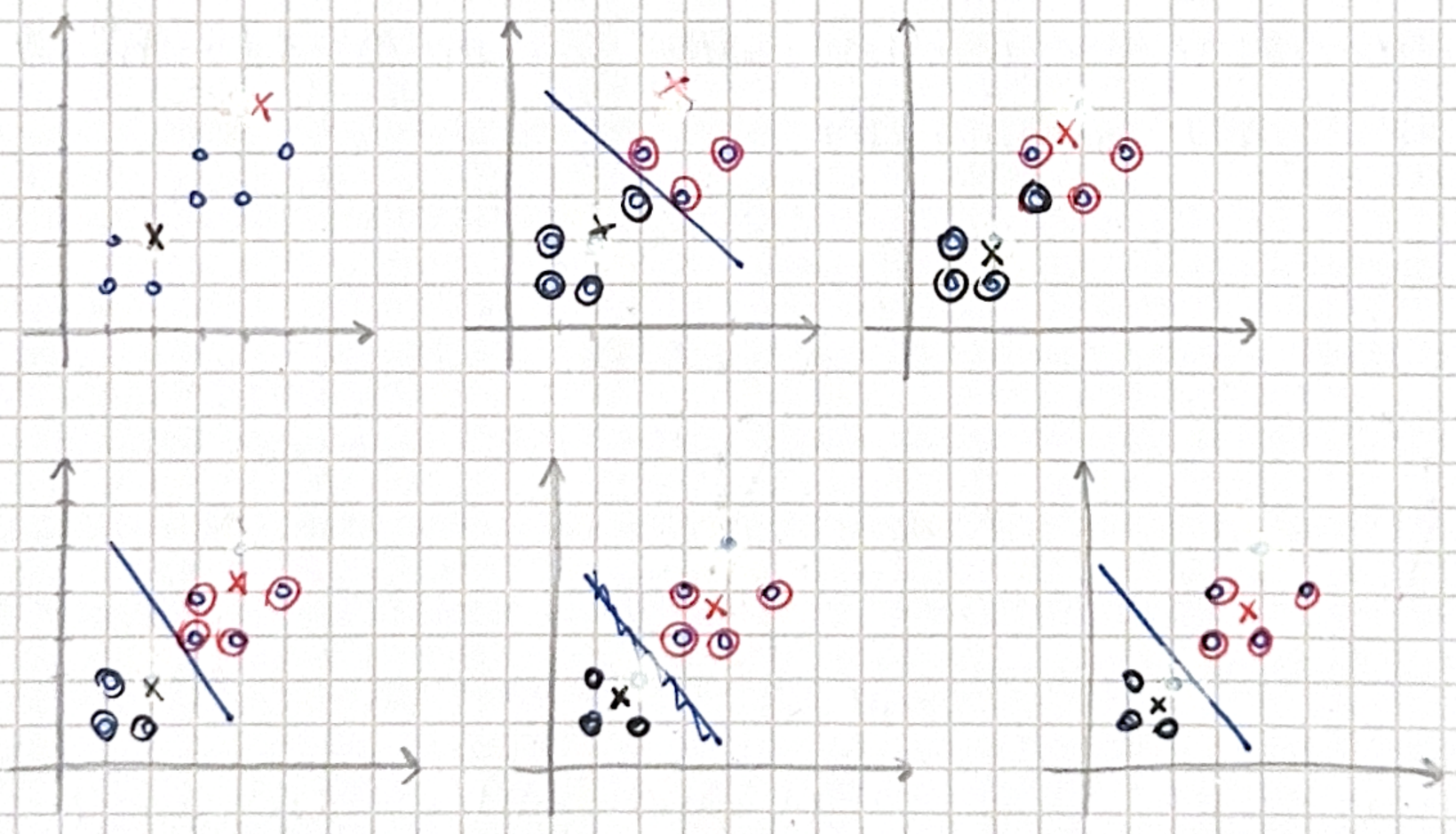
$$w_{in} = -\frac{3 + 2p + s}{s-1} \qquad h_{in} = -\frac{3 + 2p + s}{s-1}$$

$s \neq 1$ $s=2$ $P=1$

QUESTION 6

2. Hyperbolic tanyent ReLU sigmoid

3.



1. begin with decision k = number of clusters (k=3)

2 put initial partition that classifie data into k cluster this can be done randomly or systematically

2.1 take first k training samples as single cluster

2.2 the remaining N-k we will assign to nearest centroid after each assignment recompute centroide

3. take each sample in sequence and compute its distance from centroid.
   if sample is not in the closest centroid switch and update the centroid of two cluster

4. repeat all steps until convergence

II.

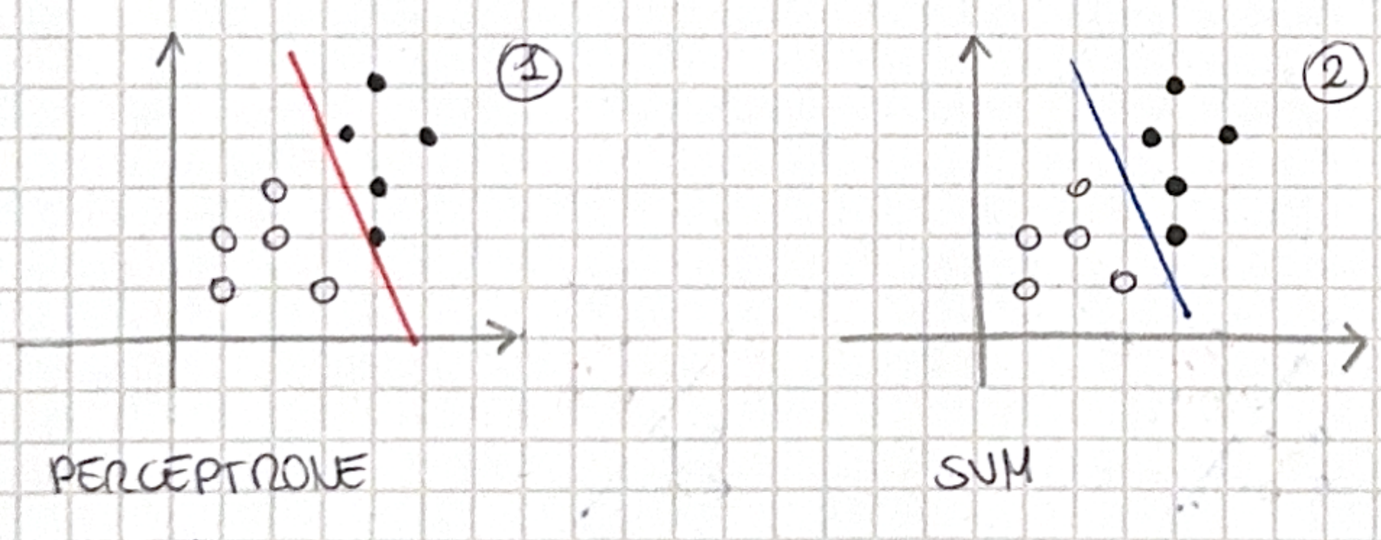
terminarion condition
1. for each switch the sum of ~~conditions~~ distance as each training samples decrease to centroid decrease
2. the are only many finitely partitions of the training examples into k cluster

ML EXAMS 12/02/2018

EXERCISE 1

supervised learning are machine problem where the model that will solve it, are trained with a dataset composed by x and y
$D = \{(x_i, y_i)\}_{i=1}^{N}$
different is the case of unsupervised learning where the model used for solve this task are trained with dataset $D = \{(x_i)\}_{i=1}^{N}$
the label are NOT used in this case

EXERCISE 2



the difference between perceptrone and SVM is perceptron is sequential algorithm ~~based~~ that depend on learning rate η ~~data~~ while SVM try to maximite the margine so in this case due the fact data is linear separable I will use SVM