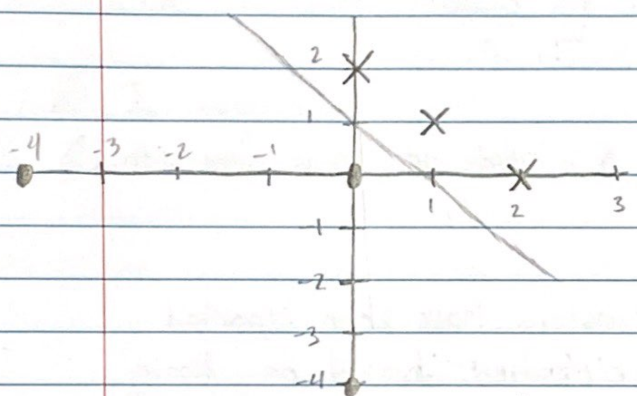
Desiderio Pilla
12/1/19
CISC 684

Homework 4
Support Vector Machine

1. $\omega_1 x_1 + \omega_2 x_2 + \omega_3 = 0$

$\omega_1^2 + \omega_2^2 + \omega_3^2 = 1$



$(0,1)$ $\quad 0 + \omega_2 + \omega_3 = 0$

$\omega_2 = -\omega_3$

$(1,0)$ $\quad \omega_1 + 0 + \omega_3 = 0$

$\omega_1 = \omega_2 = -\omega_3$

$\omega_1^2 + \omega_2^2 + \omega_3^2 = 1$

$3\omega_1^2 = 1$

$\omega_1^2 = 1/3$

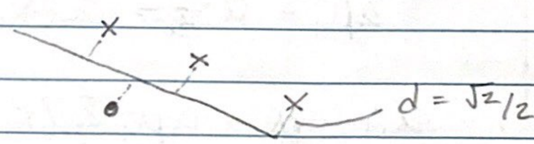
a) $\omega_1 = \omega_2 = \sqrt{3}/3$

$\omega_3 = -\sqrt{3}/3$

b) The support vectors are

$(0,0)$ $\quad (0,2)$ $\quad (1,1)$ $\quad (2,0)$

These points are all closest to the line, with a distance of $d = \sqrt{2}/2$



c) The expression for the dual problem is

$\max \sum_i^6 \alpha_i - \frac{1}{2} \sum_i^6 \sum_j^6 \alpha_i \alpha_j y_i y_j x_i^T x_j$

subject to $\alpha_i \geq 0$ and $\omega_1^2 + \omega_2^2 + \omega_3^2 = 1$ and $\sum \alpha_i y = 0$

$\sum \alpha_i^2 x_i^2 y_i^2 + b^2 = 1$

Clustering

2. $L = \sum_{j=1}^{k} \sum_{x \in S_j} (x_i - \mu_j)^2$

k clusters
$x_1, \dots, x_n$ sample points
$\mu_1, \dots, \mu_k$ centers
S_j is a set for a cluster

a) In the k-means algorithm, the update step for the center μ_1 is:

$$\mu_1 = \frac{1}{S_1} \sum_{x_i \in S_1} x_i$$ (calculate new mean of cluster 1)

Once the centers for all clusters have been updated, data points will be re-classified based on their distance to each center.

minimize $\Sigma \, dist(x, c(x))$

b) Using batch gradient descent:

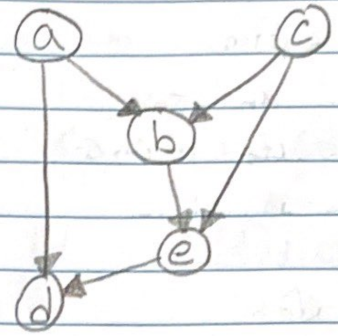
$$L = \frac{1}{2} \Sigma (x_i - \mu_1)^2$$
$$\Delta \mu_1 = \alpha \frac{\partial L}{\partial \mu_1} = \Sigma (x_i - \mu_1)(-\mu_1)$$

$$\mu_1 = \mu_1 - \alpha \mu_1 \Sigma (x_1 - \mu_1)$$

α Learning rate

Bayesian Network

3.1

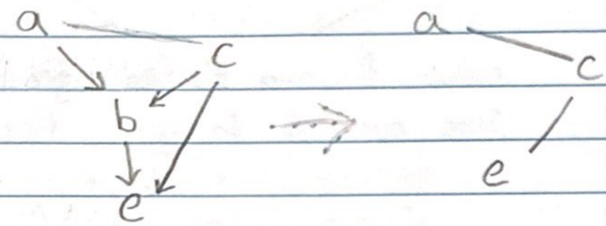


a) Factors:

- $f_1 = P[a]$
- $f_2 = P[b \mid a, c]$
- $f_3 = P[c]$
- $f_4 = P[d \mid a, e]$
- $f_5 = P[e \mid b, c]$

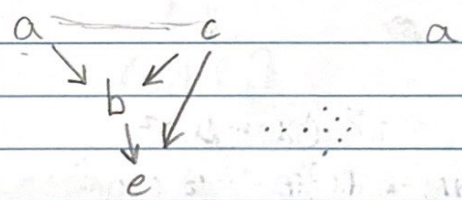
b) Independence

(a) $a \perp e \mid b$



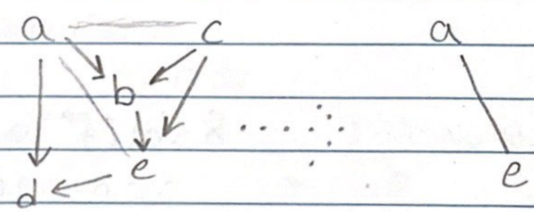
A and E are connected through C, so they are <u>not</u> necessarily independent.

(b) $a \perp e \mid b, c$



A and E are not connected, so they <u>are</u> independent.

(c) $a \perp e \mid b, c, d$



A and E are connected, so they are <u>not</u> independent.

c) see next page

3.1 c) Inferring $P[d]$ along the order (b, c, a, e)

Start with:

$f_1(a)\, f_2(a,b,c)\, f_3(c)\, f_4(a,d,e)\, f_5(b,c,e)$

b: multiply f_2 and f_5 to get $f_6(a,b,c,e)$ — $Val(f^*)$

Sum-out b to get $f_7(a,c,e)$ — $Val(f^*)$

$\downarrow$

$f_1(a)\, f_3(c)\, f_4(a,d,e)\, f_7(a,c,e)$

c: multiply f_3 and f_7 to get $f_8(a,c,e)$ — $Val(f^*)$

Sum-out c to get $f_9(a,e)$ — $Val(f^*)$

$\downarrow$

$f_1(a)\, f_4(a,d,e)\, f_9(a,e)$

a: multiply f_1 and f_4 and f_9 to get $f_{10}(a,d,e)$ — $2\,Val(f^*)$

Sum-out a to get $f_{11}(d,e)$ — $Val(f^*)$

$\downarrow$

$f_{11}(d,e)$

$\downarrow$

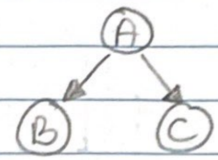
e: Sum-out e to get $f_{12}(d)$ — $Val(f^*)$

$f_{12}(d) = P[d]$

Complexity: $8\,Val(f^*)$

3.2 Distributions: $P[A]$, $P[B|A]$, $P[C|A]$

a) Draw the network



$P[A] = 0.9$

$P[B=1|A=1] = 0.1$ $\qquad$ $P[C=1|A=1] = 0.7$

$P[B=1|A=0] = 0.6$ $\qquad$ $P[C=0|A=0] = 0.3$

b)

A	B	C	Weights	
0	1	? → 1	$0.1 \times 0.6 \times 0.7 = 0.042$	?
0	1	1	$0.1 \times 0.6 \times 0.7 = 0.042$	
? → 1	0	1	$0.9 \times 0.9 \times 0.7 = 0.567$	?
1	1	? → 1	$0.9 \times 0.1 \times 0.7 = 0.063$	?
1	0	? → 1	$0.9 \times 0.9 \times 0.7 = 0.567$	?
0	0	0	$0.1 \times 0.4 \times 0.3 = 0.012$	
1	1	1	$0.9 \times 0.1 \times 0.7 = 0.063$	

c) $P[A] = \dfrac{0.567 + 0.063 + 0.567 + 0.063}{0.042 + 0.042 + 0.567 + 0.063 + 0.567 + 0.012 + 0.063} = 0.93$

$P[B=1|A=1] = \dfrac{0.063 + 0.063}{0.063 + 0.063 + 0.567 + 0.567} = 0.1$

$P[B=1|A=0] = \dfrac{0.042 + 0.042}{0.042 + 0.042 + 0.012} = 0.88$

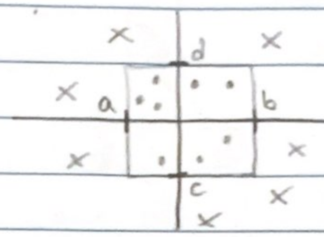
$P[C=1|A=1] = \dfrac{0.567 + 0.567 + 0.063 + 0.063}{0.567 + 0.567 + 0.063 + 0.063} = 1.0$

$P[C=0|A=0] = \dfrac{0.012}{0.012 + 0.042 + 0.042} = 0.13$

M-Step

Learning Theory

4. $H = \{(a \le x \le b) \wedge (c \le y \le d) \mid a,b,c,d \in \mathbb{R}\}$
$X = \{(x,y) \mid x,y \in \mathbb{R}\}$



Binary classification based on whether the point lies in- or outside the rectangle.

a) $VC(H) = 4$

b) With probability at least 95% output a hypothesis with error at most 0.15.

$$\epsilon = 0.15 \qquad 1-\delta = 0.95 \rightarrow \delta = 0.05$$

$$m \ge \frac{1}{\epsilon}\left[4\log\frac{2}{\delta} + 8VC(H)\log\frac{13}{\epsilon}\right]$$
$$= \frac{1}{0.15}\left[4\log\frac{2}{0.05} + 8\cdot 4\log\frac{13}{0.15}\right]$$
$$m = 1515.23$$

A training sample must have a size of 1516 or more to meet the desired results.