1. $d(A,B) = |1-1| + |4-3| = 1$
$d(A,C) = |1-3| + |4-4| = 2$
$d(A,D) = |1-5| + |4-2| = 6$
$d(A,E) = |1-3| + |4-2| = 4$
$d(A,F) = |1-3| + |4-0| = 6$
$d(B,C) = |1-3| + |3-4| = 3$
$d(B,D) = |1-5| + |3-2| = 5$
$d(B,E) = |1-3| + |3-2| = 3$
$d(B,F) = |1-3| + |3-0| = 5$
$d(C,D) = |3-5| + |4-2| = 4$
$d(C,E) = |3-3| + |4-2| = 2$
$d(C,F) = |3-3| + |4-0| = 4$
$d(D,E) = |5-3| + |2-2| = 2$
$d(D,F) = |5-3| + |2-0| = 4$
$d(E,F) = |3-3| + |2-0| = 2$

So the matrix D_1 of pairwise distances will be

	A	B	C	D	E	F
A	0	1	2	6	4	6
B	1	0	3	5	3	5
C	2	3	0	4	2	4
D	6	5	4	0	2	4
E	4	3	2	2	0	2
F	6	5	4	4	2	0

In this table, $D_1(A,B) = 1$ is the lowest value of D_1, so I cluster elements, A and B.

First Branch

Let u denote the node to which A and B are connected.
setting $\delta(A,u) = \delta(B,u) = D_1(A,B)/2 = 1/2 = 0.5$

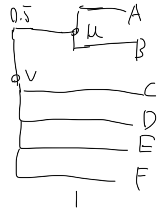
First distance matrix update

$D_2((A,B),C) = \min(D_1(A,C), D_1(B,C)) = \min(2,3) = 2$
$D_2((A,B),D) = \min(D_1(A,D), D_1(B,D)) = \min(6,5) = 5$
$D_2((A,B),E) = \min(D_1(A,E), D_1(B,E)) = \min(4,3) = 3$
$D_2((A,B),F) = \min(D_1(A,F), D_1(B,F)) = \min(6,5) = 5$

Second clustering

	(A,B)	C	D	E	F
(A,B)	0	2	5	3	5
C	2	0	4	2	4
D	5	4	0	2	4
E	3	2	2	0	2
F	5	4	4	2	0

Here $D_2((A,B),C) = 2 = D_2(C,E) = D_2(D,E) = D_2(E,F)$
Let v denote the node (A,B), C, D E, F links
Here $\delta(A,v) = \delta(B,v) = \delta(C,v) = 2/2 = 1$
$\delta(u,v) = \delta(A,v) - \delta(A,u) = 0.5$



2. (a) $Z = (1-\varepsilon)e^{-a} + \varepsilon e^{a}$

$$\frac{dZ}{da} = -(1-\varepsilon)e^{-a} + \varepsilon e^{a}$$

let the derivative form $= 0$

$$-(1-\varepsilon)e^{-a} + \varepsilon e^{a} = 0$$

$$\therefore \varepsilon e^{a} = (1-\varepsilon)e^{-a}$$

$$a = \frac{\ln(1-\varepsilon) - \ln(\varepsilon)}{2}$$

(b) First consider the case $y = -1$

then I want to prove : $\mathbb{1}(H(x) \neq -1) \leq \exp(f(x))$

$H(x) = \mathrm{sign}(f(x))$

① If $f(x) \geq 0$ then $\mathbb{1}(H(x) \neq -1) = 1$

$\exp(f(x)) \geq 1$

then $1 \leq \exp(f(x))$

$\mathbb{1}(H(x) \neq -1) \leq \exp(f(x))$

② If $f(x) < 0$, Then $\mathbb{1}(H(x) \neq -1) = 0$

$0 < \exp(f(x)) < 1$

then $0 < \exp(f(x))$

$\mathbb{1}(H(x) \neq -1) \leq \exp(f(x))$

Second consider the case $y = 1$

then I want to prove $\mathbb{1}(H(x) \neq 1) \leq \exp(-f(x))$

① If $f(x) > 0$, then $\mathbb{1}(H(x) \neq 1) = 0$

$0 < \exp(-f(x)) < 1$

then $\mathbb{1}(H(x) \neq 1) < \exp(-f(x))$

② If $f(x) \leq 0$, then $\mathbb{1}(H(x) \neq 1) = 1$,

$\exp(-f(x)) \geq 1$

$\mathbb{1}(H(x) \neq 1) \leq \exp(-f(x))$

Thus, I proved that, for all values of x and either $y = -1$ or $y = 1$,

$$\mathbb{1}(H(x) \neq y) \leq \exp(-yf(x)) = \exp\left(-y\sum_{t=1}^{T} a_t h_t(x)\right)$$

(c) $\prod_{t=1}^{t} Z_t = Z_1 \cdot Z_2 \cdot Z_3 \cdots Z_t$

$Z_1 = \sum_{i=1}^{m} D_1(x_i)\exp(-a_1 y_i h_1(x_i))$ $\quad D_1(x_i) = \frac{1}{m}$

$Z_2 = \sum_{i=1}^{m} D_2(x_i)\exp(-a_2 y_i h_2(x_i))$ $\quad D_2(x_i) = \frac{D_1(x_i)\cdot\exp(-a_1 y_i h_1(x_i))}{Z_1}$

$Z_3 = \sum_{i=1}^{m} D_3(x_i)\exp(-a_3 y_i h_3(x_i))$ $\quad D_3(x_i) = \frac{D_2(x_i)\cdot\exp(-a_2 y_i h_2(x_i))}{Z_2}$

taking D_t out of the summation,

$$Z_1 = D_1 \sum_{i=1}^{m} \exp(-y_i a_1 h_1(x_i))$$

$$Z_2 = \sum_{i=1}^{m} \frac{D_1(i)\exp(-a_1 y_i h_1(x_i))}{Z_1} \cdot \exp(-a_2 y_i h_2(x_i))$$

$$= \frac{D_1}{Z_1}\sum_{i=1}^{m} \exp(-y_i a_1 h_1(x_i) - y_i a_2 h_2(x_i))$$

$$Z_3 = \sum_{i=1}^{m} \frac{D_2(i)\exp(-a_2 y_i h_2(x_i))}{Z_2} \cdot \exp(-a_3 y_i h_3(x_i))$$

$$= \sum^{m} D_1(i)\exp(-a_1 y_i h_1(x_i)) \cdot \exp(-a_2 y_i h_2(x_i))$$

$$- \sum_{i=1} \frac{Z_1}{Z_2} \cdot \exp(-a_3 y_i h_3(x$$

$$= \sum_{i=1}^{m} \frac{D_1(i)}{Z_1 \cdot Z_2} \cdot \exp(-y_i a_1 h_1(x_i) - y_i a_2 h_2(x_i) - y_i a_3 h_3(x_i))$$

$$= \sum_{i=1}^{m} \frac{D_1}{Z_1 \cdot Z_2} \cdot \exp(-y_i \sum_{t=1}^{3} a_t h_t(x_i))$$

From here, I can see that

$$Z_T = \sum_{i=1}^{m} \frac{D_1}{Z_1 \cdot Z_2 \cdots Z_{T-1}} \cdot \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$$

$$\prod_{t=1}^{T} Z_t = Z_1 \cdot Z_2 \cdots Z_T$$
$$= \sum_{i=1}^{m} D_1 \cdot \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$$

$D_1 = \frac{1}{m}$ ↓

$$= \sum_{i=1}^{m} \frac{1}{m} \cdot \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$$

$\frac{1}{m}$ is constant, I can move it out side of $\sum_{i=1}^{m}$.

thus $\prod_{t=1}^{T} Z_t = \frac{1}{m} \sum_{i=1}^{m} \exp(-y_i \sum_{t=1}^{T} a_t h_t(x_i))$

$$\varepsilon_T = \sum_{i=1}^{m} D_T(x_i) \mathbb{I}(h_T(x_i) \neq y_i)$$

$$\mathbb{I}(h_T(x_i) \neq y_i) = 1 \Rightarrow h_T(x_i) y_i < 0 \Rightarrow \exp(-\alpha_T y_i h_T(x_i)) > 1$$

$$\mathbb{I}(h_T(x_i) \neq y_i) = 0 \Rightarrow h_T(x_i) y_i \geq 0 \Rightarrow \exp(-\alpha_T y_i h_T(x_i)) > 0$$

$\Rightarrow \mathbb{I}(h_T(x_i) \neq y_i)$ is upperbounded by $\exp(-\alpha_T y_i h_T(x_i))$

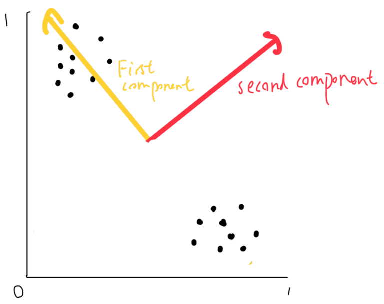
So $\varepsilon_T \leq \sum_{i=1}^{m} D_T(x_i) \exp(-\alpha_T y_i h_T(x_i)) = Z_T$

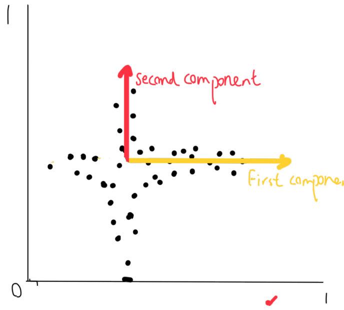
from part (a) I know that

$a = \frac{1}{2} \ln(\frac{1-\varepsilon}{\varepsilon})$ which minimize Z.

therefore it also minimize the upperbound Lost

3.(a)





(b) eigenvalue (Σ)

$$= \text{eigenvalue}\left(\begin{bmatrix} 6 & -3 & 0 \\ -3 & 6 & 0 \\ 0 & 0 & 10 \end{bmatrix} \right) = [3, 9, 10]$$

So $\sigma_1 = 10$

$\sigma_2 = 9$

$\sigma_3 = 3$

the ordered basis vector

$$[u_1, u_2, u_3] \in \mathbb{R}^3 = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{bmatrix}$$

4. (a) $L(\mu, \sigma^2; x, z) = \prod_{i=1}^{n} p(x_i \mid z_i, \mu_k) = \sum_{i=1}^{n} \sum_{k=1}^{3} \mathbb{I}[z_i = k] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_k)^2\right)$

$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right) = ($

$\mathbb{I}[z_i = k] = 1$, $\exp\left(\frac{1}{2\sigma^2}(x_i - \mu_k)^2\right)$

(b) $p(\mu_1, \mu_2, \mu_3 \mid X, Z, \gamma^2, \sigma^2)$

$= \prod_{k=1}^{3} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\gamma^2}\mu_k^2\right) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$

$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{\mu_k^2}{2\gamma^2} - \frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$

$= \prod_{k=1}^{3} \exp\left(-\frac{\mu_k}{2\gamma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$

$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{\mu_k^2}{2\gamma^2} - \frac{\mathbb{I}[z_i = k] x_i^2}{2\sigma^2} + \frac{\mathbb{I}[z_i = k] x_i \mu_k}{\sigma^2} - \frac{\mathbb{I}[z_i = k]\mu_k^2}{2\sigma^2}\right)$

$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\underbrace{\left(\frac{1}{\gamma^2} + \frac{\mathbb{I}[z_i = k]}{\sigma^2}\right)}_{a_1} \mu_k^2 + \underbrace{\frac{\mathbb{I}[z_i = k] x_i}{\sigma^2}}_{a_2} \mu_k\right)$

$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}a_1\left(\mu_k^2 - \frac{2a_2}{a_1}\mu_k + \frac{a_2^2}{a_1^2}\right)\right)$

$= \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}a_1\left(\mu_k - \frac{a_2}{a_1}\right)^2\right)$

$= \prod_{k=1}^{3} \exp\left(-\frac{1}{2\hat{\sigma}_k^2}(\mu_k - \hat{\mu}_k)^2\right)$, $\hat{\sigma}_k^2 = \left(\frac{1}{\gamma^2} + \frac{\sum \mathbb{I}(z_i = k)}{\sigma^2}\right)^{-1}$

$\propto \prod_{i=1}^{n} \exp\left(-\frac{\mu_1^2}{2\gamma^2} - \frac{1}{2\sigma^2} \mathbb{I}[z_i = 1](x_i - \mu_1)^2\right) \times \cdots \times \prod_{i=1}^{n} \exp\left(\frac{-\mu_3^2}{2\gamma^2} \cdots (x_i - \mu_3)^2\right)$

$= p(\mu_1 \mid X, Z, \gamma^2, \sigma^2) \times \cdots \times p(\mu_3 \mid X, Z, \gamma^2, \sigma^2)$

$\Downarrow$

μ_1, μ_2, μ_3 are conditionally independent

(c) $p(Z \mid \mu, X, \gamma^2, \sigma^2) \propto \prod_{k=1}^{3} \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2} \mathbb{I}[z_i = k](x_i - \mu_k)^2\right)$

$\propto \prod_{k=1}^{3} \prod_{i=1}^{n} \left[\exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma^2}\right)\right]^{\mathbb{I}[z_i = k]}$

$= \prod_{i=1}^{n} \exp\left(-\frac{(x_i - \mu_1)^2}{2\sigma^2}\right)^{\mathbb{I}[z_i = 1]} \cdot \exp\left(-\frac{x_i - \mu_2}{2\sigma^2}\right)^{\mathbb{I}[z_i = 2]} \cdots$

Thus Z is not independent given μ_1, μ_2, μ_3