Homework 8 Solutions

#11 $\quad W \in \mathbb{R}^{K\times D}, \quad x \in \mathbb{R}^D \qquad f(x,W) = \text{Softmax}(W\underline{x})$

a.) $$f_c(\underline{x}, W) = \frac{\exp(w_c^T x)}{\sum_k \exp(w_k^T x)}$$

w_k^T = k-th row of W
f_c = c-th component of f

b.) $$f_c(\underline{x}, W+A) = \frac{\exp((w_c^T + a)x)}{\sum_{c'} \exp((w_{c'}+a)^T x)} = \frac{\exp(w_c^T x)\cdot \cancel{\exp(a^T x)}}{\sum_{c'} \exp(w_{c'}^T x)\cdot \cancel{\exp(a^T x)}} = f_c(x,W)$$

c.) i.) Set $\underline{a} = W\underline{x}$, then

$$a_m = w_m^T x = \sum_d w_{md} x_d \Rightarrow \begin{cases} \dfrac{\partial a_m}{\partial w_{kj}} = \sum_d \delta_{km}\delta_{jd} x_d = \delta_{km} x_j \\ \dfrac{\partial a_m}{\partial x_j} = \sum_d w_{md}\delta_{dj} = w_{mj} \end{cases}$$

ii.) Set $$S_c(\underline{a}) = \frac{\exp(a_c)}{\sum_{c'} \exp(a_{c'})}$$ (c-th component of softmax fcn), then

$$\frac{\partial S_c}{\partial a_m} = \frac{\frac{\partial}{\partial a_m}\exp(a_c)}{\sum_{c'}\exp(a_{c'})} - \frac{\exp(a_c)}{\left[\sum_{c'}\exp(a_{c'})\right]^2}\cdot\frac{\partial}{\partial a_m}\sum_{c'}\exp(a_{c'})$$

$$= \frac{\delta_{mc}\cdot\exp(a_c)}{\sum_{c'}\exp(a_{c'})} - \frac{\exp(a_c)\cdot\exp(a_m)}{\left[\sum_{c'}\exp(a_{c'})\right]^2} = \delta_{mc} S_c(a) - S_c(a)\cdot S_m(a)$$

with this, compute the partial derivatives:

$$\frac{\partial}{\partial w_{kj}} f_c(\underline{x}, W) \overset{a=Wx}{=} \frac{\partial}{\partial w_{kj}} S_c(a) = \sum_m \frac{\partial S_c}{\partial a_m}\cdot\frac{\partial a_m}{\partial w_{kj}}$$

$$= \sum_m \left[\delta_{mc} S_c(a) - S_c(a)\cdot S_m(a)\right]\cdot\delta_{km} x_j = \delta_{kc} S_c(a) x_j - S_c(a) S_k(a) x_j$$

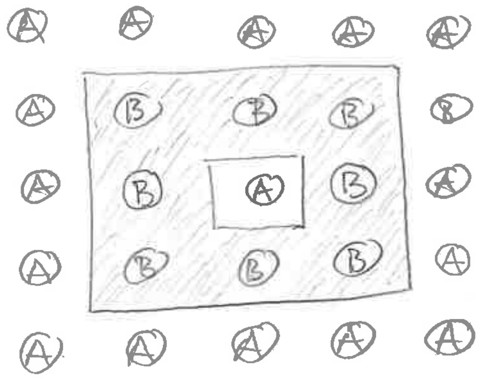
$$= S_c(a) x_j \left[\delta_{kc} - S_k(a)\right]$$

$$\frac{\partial}{\partial x_j} f_c(x,W) = \frac{\partial}{\partial x_j} S_c(a) = \sum_m \frac{\partial S_c}{\partial a_m}\cdot\frac{\partial a_m}{\partial x_j}$$

$$= \sum_m \left[\delta_{mc} S_c(a) - S_c(a)\cdot S_m(a)\right]\cdot w_{mj} = S_c(a)\cdot\left[w_{cj} - \sum_m w_{mj} S_m(a)\right]$$

HW 8 #2 |

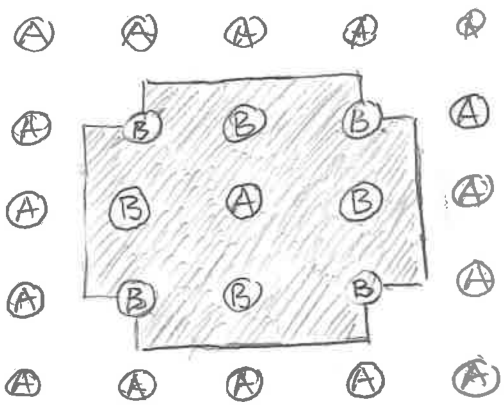
K=1



→ B

rest → A

K=3



The point (2,2) is occupied by A but is classified by B b/c there are more B-neighbors

HW8 #3 | (n_1, n_2, n_3) where $n_1 \geq n_2 \geq n_3$ means:

the largest cluster contains n_1 data points
the 2nd largest cluster contains n_2 data points
the 3rd largest cluster contains n_3 data points

Cases

a) $(5, 0, 0)$: all data are in the same cluster $\rightarrow$ 1 possibility

b) $(4, 1, 0)$: $\binom{5}{4} = 5$ possibilities for the largest then the other clusters are fixed $\rightarrow$ 5 possibilities

c) $(3, 2, 0)$: $\binom{5}{3} = 10$ possibilities for the largest then the other clusters are fixed $\rightarrow$ 10 possibilities

e) $(2, 2, 1)$: $\binom{5}{2} = 10$ possibilities for the first
$\binom{3}{2} = 3$ possibilities for the second $\rightarrow$ 30 possibilities
then the third cluster is fixed

d) $(3, 1, 1)$: $\binom{5}{3} = 10$ possibilities for the largest, then the other two are fixed $\rightarrow$ 10 possibilities

56 total

when counting these keep in mind that
- data points are distinct
- clusters don't have distinct numbers

so, for instance,

$x_1\ x_3\ x_5$ $\quad$ x_4 $\quad$ x_2

counts as the same clustering

$x_5\ x_3\ x_1$ $\quad$ x_2 $\quad$ x_4

HW8 #4

(n_1, n_2) means: n_1 data points are in the bigger cluster
n_2 data points are in the smaller cluster

case 1: N is even

$$\left.\begin{matrix} (N,0) \\ (N-1,1) \\ \vdots \\ \left(\frac{N}{2}, \frac{N}{2}\right) \end{matrix}\right\}$$

for $(N-k, k)$ we have $\binom{N}{k}$ possibilities to place x_i's into the smaller cluster. In each case the bigger cluster is fixed.

$\Rightarrow$ total number of clusterings $= \sum_{k=0}^{N/2} \binom{N}{k}$

case 2: N is odd

$$\left.\begin{matrix} (N,0) \\ (N-1,1) \\ \vdots \\ \left(\frac{N+1}{2}, \frac{N-1}{2}\right) \end{matrix}\right\}$$

Same argument as before

total number of clusterings $= \sum_{k=0}^{\frac{N-1}{2}} \binom{N}{k}$

HW8 #5

$$D = \begin{array}{c} \\ \left[\begin{array}{cccc} 0.3 & 0 & 0.5 & 0.8 \\ 0.4 & 0.5 & 0 & 0.45 \\ 0.7 & 0.8 & 0.45 & 0 \end{array}\right] \end{array} \begin{array}{c} 2 \\ 3 \\ 4 \end{array} \qquad C_1 = \{1\} \quad C_2 = \{2\} \quad C_3 = \{3\} \quad C_4 = \{4\}$$

a.) $\underset{(i,j)}{\operatorname{argmin}}\, d_{ij} = (1,2) \Rightarrow$ merge C_1 and C_2 $\Rightarrow C_{12} = \{1,2\}$

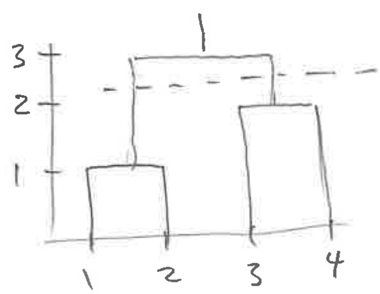
new distances $d(C_1, C_3) = \max\{d_{13}, d_{23}\} = 0.5$ (complete linkage)

$d(C_1, C_4) = \max\{d_{14}, d_{24}\} = 0.8$

$$D = \begin{array}{c} \begin{array}{ccc} (1,2) & 3 & 4 \end{array} \\ \left[\begin{array}{ccc} 0 & 0.5 & 0.8 \\ 0.5 & 0 & 0.45 \\ 0.8 & 0.45 & 0 \end{array}\right] \end{array} \begin{array}{c} \\ (1,2) \\ 3 \\ 4 \end{array}$$

merge C_3 and C_4 $\quad C_3 := \{3.4\}$

$\Rightarrow$ dendrogram



b.) $\operatorname{argmin} d_{ij} = (1,2) \Rightarrow$ merge C_1 and C_2 $\Rightarrow C_{12} = \{1,2\}$

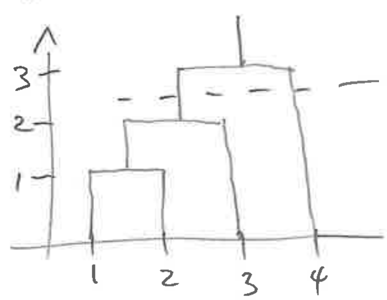
new distances $d(C_1, C_3) = \min\{d_{13}, d_{23}\} = 0.4$ (Single linkage)

$d(C_1, C_4) = \min\{d_{14}, d_{24}\} = 0.7$

$$D = \begin{array}{c|ccc} & (1,2) & 3 & 4 \\ \hline (1,2) & 0 & 0.4 & 0.7 \\ 3 & 0.4 & 0 & 0.45 \\ 4 & 0.7 & 0.45 & 0 \end{array}$$

$\Rightarrow$ merge C_1 and C_3 $\Rightarrow C_{123} = \{1, 2, 3\}$

$\Rightarrow$ dendrogram



c.) $\{1, 2\}$ and $\{3, 4\}$

d.) $\{1, 2, 3\}$ and $\{4\}$