DATE

Q1 In case of Ncut:

$$\text{Ncut}(A_1, \dots, A_K) = \frac{1}{2} \sum_{k=1}^{K} \frac{w(A_k, \bar{A}_k)}{\text{vol}(A_k)}$$

where $\text{vol}(A) = \sum_{i \in A} \sum_{j \in V} w_{i,j}$

& V is the set of all vertices in the graph.

In case $K = 2$, then

$$\text{Ncut}(A, \bar{A}) = \frac{1}{2} \left[\frac{\text{cut}(A, \bar{A})}{\text{Vol}(A)} + \frac{\text{cut}(\bar{A}, A)}{\text{Vol}(\bar{A})} \right]$$

Let $f_A = (f_{A_1}, f_{A_2}, \dots f_{A_n})^T$ where

$$f_{A_i} = \begin{cases} \sqrt{\text{vol}(\bar{A}) / \text{vol}(A)} & \text{if } i \in A \\ -\sqrt{\text{vol}(A) / \text{vol}(\bar{A})} & \text{if } i \notin A \text{ ie } i \in \bar{A} \end{cases}$$

Then $f_A^T L f_A = \frac{1}{2} \sum w_{ij} (f_{A_i} - f_{A_j})^2$

$$= \frac{1}{2} \left(\sum_{i \in A, j \in \bar{A}} w_{ij} \left(\sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}} + \sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}} \right)^2 \right)$$

$$+ \frac{1}{2} \left(\sum_{i \in \bar{A}, j \in A} w_{ij} \left(-\sqrt{\frac{\text{vol}(A)}{\text{vol}(\bar{A})}} - \sqrt{\frac{\text{vol}(\bar{A})}{\text{vol}(A)}} \right)^2 \right)$$

$$= \frac{1}{2} \text{cut}(A, \bar{A}) \left(\frac{\text{vol}(\bar{A})}{\text{vol}(A)} + \frac{\text{vol}(A)}{\text{vol}(\bar{A})} + 2 \right)$$

$$+ \frac{1}{2} \text{cut}(A, \bar{A}) \left(\frac{\text{vol}(A)}{\text{vol}(\bar{A})} + \frac{\text{vol}(\bar{A})}{\text{vol}(A)} + 2 \right)$$

$$= \text{cut}(A, \bar{A}) \left(\frac{\text{vol}(A)}{\text{vol}(\bar{A})} + \frac{\text{vol}(\bar{A})}{\text{vol}(A)} + 2 \right)$$

$$\begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} \begin{bmatrix} \mathrm{vol}(A)/\mathrm{vol}(\bar{A}) \\ \vdots \end{bmatrix}$$

DATE

$$= \mathrm{cut}(A, \bar{A}) \times \mathrm{vol}(V) \left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(\bar{A})} \right)$$

$$= \mathrm{vol}(V) \left(\frac{\mathrm{cut}(A, \bar{A})}{\mathrm{vol}(A)} + \frac{\mathrm{cut}(A, \bar{A})}{\mathrm{vol}(\bar{A})} \right)$$

$$= \mathrm{vol}(V) \cdot \text{Normalised Cut}(A, \bar{A})$$

&

$$\mathbb{1}^T D f_A = \sum_{i \in A} d_i \sqrt{\frac{\mathrm{vol}(\bar{A})}{\mathrm{vol}(A)}} - \sum_{i \notin A} d_i \sqrt{\frac{\mathrm{vol}(A)}{\mathrm{vol}(\bar{A})}}$$

where $d_i = \sum_{j \in V} w_{ij}$

$$\therefore \quad = \sqrt{\frac{\mathrm{vol}(\bar{A})}{\mathrm{vol}(A)}} \times \sum_{i \in A} \sum_{j \in V} w_{ij} - \sum_{i \in \bar{A}} \sum_{j \in V} w_{ij} \times \sqrt{\frac{\mathrm{vol}(A)}{\mathrm{vol}(\bar{A})}}$$

& $\mathrm{vol}(A) = \sum_{i \in A} \sum_{j \in V} w_{ij}$

$$= \sqrt{\frac{\mathrm{vol}(\bar{A})}{\mathrm{vol}(A)}} \times \mathrm{vol}(A) - \sqrt{\frac{\mathrm{vol}(A)}{\mathrm{vol}(\bar{A})}} \times \mathrm{vol}(\bar{A})$$

$$= \sqrt{\mathrm{vol}(\bar{A}) \cdot \mathrm{vol}(A)} - \sqrt{\mathrm{vol}(A) \cdot \mathrm{vol}(\bar{A})}$$

$$= 0$$

$$f_A^T D f_A = \frac{\mathrm{vol}(\bar{A})}{\mathrm{vol}(A)} \sum_{i \in A} d_i + \frac{\mathrm{vol}(A)}{\mathrm{vol}(\bar{A})} \sum_{i \notin A} d_i$$

$$= \frac{\mathrm{vol}(\bar{A})}{\cancel{\mathrm{vol}(A)}} \times \cancel{\mathrm{vol}(A)} + \frac{\mathrm{vol}(A)}{\cancel{\mathrm{vol}(\bar{A})}} \times \cancel{\mathrm{vol}(\bar{A})}$$

Typo

DATE

$$= \sum_{i \in A} \sum_{j \in V} w_{ij} + \sum_{i \notin A} \sum_{j \in V} w_{ij}$$

$$= \sum_{i \in V} \sum_{j \in V} w_{ij} = vol(V)$$

$\therefore$ We can re write the optimisation problem as

$$\min_{A \subset \{1, \dots n\}} f_A^T L f_A$$

s.t

$$\underline{1}^T D f_A = 0$$

$$\& \ \cancel{\|f_A\|} \ f_A^T D f_A = vol(V)$$

Now, a relaxation of above problem is

$$\min_{f \in \mathbb{R}^n} f^T L f \quad — (1)$$

$$\text{s.t} \quad \underline{1}^T D f_A = 0$$
$$\& \quad f_A^T D f_A = vol(V)$$

Let $g = D^{1/2} f$

$\therefore f = D^{-1/2} g.$

$\&$ $f^T = g^T D^{-1/2}$ {as $D = D^T$ {diagonal matrix}}

Subsituting f^T in eq (1) we get.

$$\min_{g \in \mathbb{R}^n} g^T D^{-1/2} L D^{-1/2} g$$

$$\text{s.t } \underline{1}^T D^{1/2} g = 0$$
$$\& \ \cancel{g_A^T} \ \cancel{g_A} = vol(V)$$

Typo

ii Putting $GR(g)$, g is the second smallest eigenvector of $D^{-1/2} L D^{-1/2}$.

Let $D^{-1/2} L D^{-1/2} = U \Lambda U^T$

or $D^{-1/2} L D^{-1/2} u_2 = \lambda_2 u_2$.

& $u_2 = g$.

$\therefore D^{-1/2} L D^{-1/2} g = \lambda_2 u_2$

~~or g~~ $= D^{-1/2} L D^{-1/2} g = \lambda_2 g$

$D^{-1/2} L f = \lambda_2 D^{1/2} f$

$\Rightarrow \quad D^{-1} L f = \lambda_2 f$

ie f is the second smallest eigenvector of $D^{-1} L$.

DATE

Q2

Training Data : $(x^{(i)}, y^{i})$, $i=1 \ldots n$

hidden layer = 1

$$\sigma(t) = 1/(1+e^{-t})$$

$$z_{ij} = \sigma(a_{ij})$$

$$a_{ij} = \sum_{r=1}^{5} \sum_{s=1}^{5} w^{(0)}_{rs} \, x_{i+r-1,\, j+s-1}$$

$$y_k = \sum_{j=1}^{124} \sum_{i=1}^{124} w^{(1)}_{ij} \, z_{ij}$$

$$\epsilon = \sum \Big(y_i - \overbrace{f(x_i)}^{\hat{y}}\Big)^2$$

$$\frac{\partial \epsilon}{\partial w} = \frac{\partial \epsilon}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w}$$

$$\frac{\partial \epsilon}{\partial w^{(0)}} = \frac{\partial \epsilon}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial a} \frac{\partial a}{\partial w^{(0)}}$$

Let $\epsilon = \sum_{k=1}^{n} (y_k - \underbrace{f(x_k)}_{\hat{y}_k})^2 = \sum_{k=1}^{n} \epsilon_k$

$$\frac{\partial \epsilon}{\partial w^{(1)}_{ij}} = \sum \frac{\partial \epsilon_k}{\partial \hat{y}_k} \times \frac{\partial \hat{y}_k}{\partial w^{(1)}_{ij}}$$

$$= \sum_{k=1}^{n} 2 . (y_k - \hat{y}_k) . (-1) . z_{ij}$$

$$\frac{\partial \epsilon}{\partial w^{(1)}_{ij}} = -2 z_{ij} \sum_{k=1}^{n} (y_k - \hat{y}_k)$$

||ly $\frac{\partial \epsilon}{\partial w^{(0)}_{rs}} = \sum \frac{\partial \epsilon_k}{\partial \hat{y}_k} \cdot \frac{\partial \hat{y}_k}{\partial z_{ij}} \frac{\partial z_{ij}}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial w^{(0)}_{rs}}$

$\frac{\partial \epsilon_k}{\partial \hat{y}_k} = -2(y_k - \hat{y}_k)$; $\frac{\partial \hat{y}_k}{\partial z_{ij}} = w^{(1)}_{ij}$

$t - 4 \leq p < t$

DATE

$$\frac{\partial E}{\partial w_{pq}^{(0)}} = \sum_{k=1}^{n} \frac{\partial E_k}{\partial \hat{y}_k} \sum_{u=1}^{124} \sum_{t=1}^{124} \frac{\partial \hat{y}_k}{\partial z_{tu}} \cdot \frac{\partial z_{tu}}{\partial a_{tu}} \cdot \frac{\partial a_{tu}}{\partial w_{pq}^{(0)}}$$

$$\frac{\partial E_k}{\partial \hat{y}_k} = -2(y_k - \hat{y}_k) \quad ; \quad \frac{\partial \hat{y}_k}{\partial z_{tu}} = w_{tu}^{(1)}$$

$$\frac{\partial z_{tu}}{\partial a_{tu}} = \sigma'(a_{tu}) = \frac{1}{(1+e^{-a_{tu}})^2} \cdot e^{-a_{tu}}$$

$$\frac{\partial a_{tu}}{\partial w_{pq}^{(0)}} = \cancel{w_{pq}} \; x_{t+p-1,\, u+q-1}$$

$$\cancel{\frac{\partial E}{\partial w_{pq}^{(0)}} = \sum_{u=q}^{q+4} \sum_{k=1}^{n} \sum_{t=p}^{p+4} -2(y_k - \hat{y}_k) \cdot w_{tu}^{(1)} \frac{e^{-a_{tu}}}{(1+e^{-a_{tu}})^2} \cdot x_{t+p-1,\, u+q-1}}$$

$$\frac{\partial E}{\partial w_{pq}^{(0)}} = \sum_{u=q}^{q+4} \sum_{k=1}^{n} \sum_{t=p}^{p+4} -2(y_k - \hat{y}_k)\, w_{tu}^{(1)} \frac{e^{-a_{tu}}}{(1+e^{-a_{tu}})^2} \cdot x_{t+p-1,\, u+q-1}$$

$$\frac{\partial E}{\partial w_{pq}^{(0)}} = \sum_{k=1}^{n} \sum_{u=1}^{124} \sum_{t=1}^{124} (-2)(y_k - \hat{y}_k)\, w_{tu}^{(1)} \frac{e^{-a_{tu}}}{(1+e^{-a_{tu}})^2} \cdot x_{t+p-1,\, u+q-1}$$

Typo

$$Q.\ \cancel{J(f)} = \min_{f \in \mathcal{H}} \lambda \|f\|_{\mathcal{H}}^2 + \frac{1}{m} \sum_{i=1}^{m} (y_i - f(x_i))^2 + \gamma \frac{1}{2} \sum_{i,j=1}^{m+n} w_{ij} (f(x_i) - f(x_j))^2$$

- ~~ie $\min_{f \in \mathcal{H}} J$~~ ie $\min_{f \in \mathcal{H}} J(f)$

$$J(f) = \underbrace{\lambda \|f\|_{\mathcal{H}}^2}_{\Lambda(\|f\|^2)} + \underbrace{\frac{1}{m} \sum_{i=1}^{m} (y_i - f(x_i))^2 + \gamma \frac{1}{2} \sum_{i,j=1}^{m+n} w_{ij} (f(x_i) - f(x_j))^2}_{L(f_1, f_2, \ldots, f_{m+n})}$$

~~ab~~ & Λ is ~~non decreasing~~ strictly increasing.
as $\lambda > 0$

$\therefore$ According to representer theorem, we can say that the solution will have the following form.

$$f = \sum_{i=1}^{m+n} \alpha_i k(\cdot\,; x_i)$$

$$\text{or } f(x) = \sum_{i=1}^{m+n} \alpha_i k(x, x_i)$$

$\therefore$ Putting this value of f in the optimisation problem, we get.

$$\lambda \|f\|_{\mathcal{H}}^2 = \lambda \langle \sum \alpha_i k(\cdot\,; x_i); \sum \alpha_i k(\cdot\,; x_i) \rangle$$

$$= \lambda \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j)$$

$$= \lambda \alpha^T K \alpha$$

where $\alpha = [\alpha_1 \ldots \alpha_m \alpha_{m+1} \ldots \alpha_{m+n}]$

& $K = [k(x_i, x_j)]_{i,j}$ $(m+n) \times (m+n)$

DATE

$$\frac{1}{m}\sum_{i=1}^{m}(y_i - f(x_i))^2 = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{m+n}\alpha_j k(x_i, x_j)\right)^2$$

$$= \frac{1}{m}\|y - K_m\alpha\|^2$$

$$\frac{\gamma}{2}\sum_{i,j=1}^{m+n} w_{ij}\,(f(x_i) - f(x_j))^2 = \frac{\gamma}{2} f'^T L' f'$$

where $L' = D - W$ {as defined for spectral clustering}

$$\& \; f' = \begin{bmatrix} \sum_{i=1}^{m+n} k(x_i, x_1)\,\alpha_i \\ \vdots \\ \sum_{i=1}^{m+n} k(x_i, x_{m+n})\alpha_i \end{bmatrix}$$

$$= K_{m+n}\,\alpha$$

$$\therefore\; f'^T L' f' = \frac{\gamma}{2}\left(\alpha^T K_{m+n}^T L' K_{m+n}\,\alpha\right)$$

$$\therefore\; \min_{f\in\mathcal{H}} J(f) = \lambda\,\alpha^T K\alpha + \frac{1}{m}\|Y - K_m\alpha\|^2 + \frac{\gamma}{2}\left(\alpha^T K_{m+n}^T L' K_{m+n}\,\alpha\right)$$

$$\therefore\; = \lambda\alpha^T K\alpha + \frac{1}{m}\left(y^T y - 2y^T K_m\alpha + \alpha^T K_m^2\alpha\right) + \frac{\gamma}{2}\left(\alpha^T K_{m+n}^T L' K_{m+n}\,\alpha\right)$$

Differentiating w.r.t α we get.

$$= 2\lambda K\alpha + \frac{1}{m}\left(-2\,K_m y + 2K_m^2\alpha\right) + \frac{\gamma}{2}\times 2\left(K^T L' K\right)\alpha = 0$$

$$= \lambda K\alpha - \frac{1}{m}K_m y + \frac{1}{m}K_m^2\alpha + \frac{\gamma}{2}K^T L' K\alpha = 0$$

Typo

Cwt

$$\Rightarrow \left(\lambda K + \frac{K_m^2}{m} + \frac{\gamma}{\circledcirc} K^T L' K\right)\alpha^{\circ} = \frac{1}{m} K_m y$$

$$\text{or } \alpha = \frac{1}{m}\left(\lambda K + \frac{K_m^2}{m} + \frac{\gamma}{\circledcirc} K^T L' K\right)^{-1} K_m y$$

where $K_{\circ} = \begin{bmatrix} \boxed{K_m} \\ \boxed{K_n} \end{bmatrix}$

Q5 When $\phi(t) = e^{-t}$, then

$$\alpha_t = \arg\min_{\alpha} \frac{1}{n} \underbrace{\sum_{i=1}^{n} \phi\left(y_i F_{t-1}(x_i) + y_i \alpha f_t(x_i)\right)}_{J}$$

$\because$ We are using exponential loss here

$\therefore$

$$\therefore J = \frac{1}{n} \sum_{i=1}^{n} e^{-y_i F_{t-1}(x_i)} \cdot e^{-y_i \alpha f_t(x_i)} \quad -(1)$$

Let us look at w_i^{t+1}

$$w_i^{t+1} = \frac{\phi'(y_i F_t(x_i))}{\underbrace{\sum_{j=1}^{n} \phi'(y_j F_t(x_j))}} \longrightarrow Z_t \{ \text{normalising constant}$$

$$\text{Then } w_i^{t+1} = \frac{\phi'(y_i F_t(x_i))}{Z_t}$$

$$\phi'(y_i F_t(x_i)) = e^{-y_i F_t(x_i)}$$
$$= e^{-y_i \{F_{t-1}(x_i) + \alpha_t f_t)\}}$$
$$= e^{-y_i F_{t-1}(x_i)} \cdot e^{-y_i \alpha_t f_t}$$

$$\& \ w_i^{t} = \frac{e^{-y_i F_{t-1}(x_i)}}{Z'_{t-1}}$$

Typo

DATE

$$\therefore \quad e^{-y_i F_{t-1}(x_i)} = \omega_i^t \cdot Z_{t-1}$$

$$\therefore \quad \omega_i^{t+1} = \frac{\omega_i^t \cdot e^{-y_i \alpha_t f_t}}{Z_t} \cdot Z_{t-1}$$

$$= \omega_i^t \frac{\exp(-\alpha_t y_i f_t(x_i))}{\underbrace{K_t}}$$

$\rightarrow$ is the normalising constant $= Z_t / Z_{t-1}$

Which is similar in form to adaboost algo.

Now let us look at eq. ①, we get.

$$J = \frac{1}{n} \sum_{i=1}^{n} e^{-y_i F_{t-1}(x_i)} \cdot e^{-y_i \alpha f_t(x_i)}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \omega_i^t \; e^{-y_i \alpha f_t(x_i)}$$

$$\cancel{= \frac{1}{n} \sum_{i=1}^{n}}$$

Let us ~~define~~ break this summation into two part, one where $y_i = f_t(x_i)$ & one where they are not equal.

$$\text{ie} \quad J = \frac{1}{n} \left(\sum_{y_i = f_t(x_i)} \omega_i^t \, e^{-y_i \alpha f_t(x_i)} + \sum_{y_i \neq f(x_i)} \omega_i^{(t)} e^{\alpha f_t} \right)$$

$$\cancel{= \frac{1}{n} \left(\sum_{i=1}^{n} \omega_i^t e^{-y_i \alpha f_t} + \sum_{y_i \neq f_t(x_i)} \omega_i^t (e^{-y_i \alpha f_t} \right.}$$

$\because \; y_i = \{+1, -1\}$.

$$\therefore \; J = \frac{1}{n} \sum_{i=1}^{n} \omega_i^t e^{-\alpha} + \sum_{y_i \neq f(x_i)} \omega_i^{(t)} \left(e^{\alpha} - e^{-\alpha} \right)$$

DATE

Now

$$\frac{\partial J}{\partial \alpha} = -\frac{1}{n}\left(\sum_{i=1}^{n} w_i^t e^{-\alpha} + \sum_{y_i \neq f(x_i)} w_i^t \left(e^{\alpha} + e^{-\alpha}\right)\right)$$

$$0 = \frac{1}{n}\left(\sum_{y_i \neq f(x_i)} w_i^t e^{\alpha} - \sum_{y_i = f(x_i)} w_i^t e^{-\alpha}\right)$$

$$\text{or} \quad e^{2\alpha} = \frac{\left.\sum_{y_i = f(x_i)} w_i^t \right\} \longrightarrow a}{\left.\sum_{y_i \neq f(x_i)} w_i^t \right\} \longrightarrow b}$$

$$\text{or} \quad \alpha_{m} = \frac{1}{2} \ln\left(\frac{a}{b}\right)$$

$$\text{Let } r_t = \sum_{i=1}^{n} w_i^{(t)} \mathbb{1}_{\{f_t(x_i) \neq y_i\}}$$

$$= \sum_{f_t(x_i) \neq y_i} w_i^t$$

& we know, $\sum_{i=1}^{n} w_i^t = 1$

$$\therefore \sum_{y_i = f(x_i)} w_i^t = 1 - r_t.$$

$$\therefore \alpha_{min} = \frac{1}{2} \ln\left(\frac{1 - r_t}{r_t}\right)$$

$$\text{or} \quad \alpha_t = \frac{1}{2} \ln\left(\frac{1 - r_t}{r_t}\right)$$

~~Kly~~,

$$\& \quad F_T(x) = F_{t-1} + \alpha_t f_t = F_{t-2} + \alpha_{t-1} f_{t-1} + \alpha_t f_t$$

Typo

DATE

$$\therefore F_T(x) = \sum_{t=1}^{T} \alpha_t f_t(x)$$

$$\therefore h_T(x) = \text{sign}(F_T(x))$$

& hence its is proved.

Q4 Let n be the total number of data points ~~reaching root node T, then~~.

$$\therefore J(T) = \sum_{i=1}^{n} l(y_i, T(x_i)) + \lambda|T|$$

Let out of n data points, m data points enter T_1

$\therefore$ $m-n$ data points will enter T_2.

$$\therefore J(T_1) = \sum_{i=1}^{m} l(y_i, T_1(x_i)) + \lambda|T_1|$$

$$\& \ J(T_2) = \sum_{i=m+1}^{n} l(y_i, T_2(x_i)) + \lambda|T_2|$$

Now, we know

$$|T| = |T_1| + |T_2| \quad \text{— } 1$$

Also, the loss is defined as # errors $\therefore$

$$J(T) = \sum_{i=1}^{n} 1_{\{y_i \neq T(x_i)\}} + \lambda|T|$$

||ly

$$J(T_1) = \sum_{i=1}^{m} 1_{\{y_i \neq T(x_i)\}} + \lambda|T_1|$$

$$J(T_2) = \sum_{i=m+1}^{n} 1_{\{y_i \neq T(x_i)\}} + \lambda|T_2|$$

& $\because$ errors are additive here, we get.

$$\sum_{i=1}^{m} 1_{\{y_i \neq T_1(x_i)\}} + \sum_{i=m+1}^{n} 1_{\{y_i \neq T_2(x_i)\}} = \sum_{i=1}^{n} 1_{\{y_i \neq T(x_i)\}} \quad \text{— } 2$$

$$\therefore J(T) = J(T_1) + J(T_2)$$

Typo

$\therefore$ Using eq ① & ②,

$$J(T_1) + J(T_2) = J(T)$$

High level pseudocode for decision tree pruning:

① Start at the leaves of tree (T) and then work towards the root and apply below algorithm to each decision node (m).

Algo (T, m):

(a) Calculate $J(T)$

(b) Let $T_{smaller}$ be tree after pruning that decision node ~~below~~ m.

(c) Calculate $(T_{smaller})$

(d) If $J(T_{smaller}) < J(T)$, prune T to $T_{smaller}$.

Q3 Let $\|\omega\|_1 = S^T\omega = \omega^T S$

Where $S_i = \text{sign}(\omega_i)$ if $\omega_i \neq 0$
$= 0$ if $\omega_i = 0$

$\therefore \; g(\omega) = S$