2. (a) $\text{softmax}(\text{sim}(z_{it}, z_{jt'})/\tau)$

$$= \text{softmax}\left(\frac{z_{it}\cdot z_{jt'}}{\|z_{it}\|_2 \|z_{jt'}\|_2\cdot\tau}\right) = \frac{\exp\left(\frac{z_{it}\cdot z_{jt'}}{\|z_{it}\|_2 \|z_{jt'}\|_2\cdot\tau}\right)}{\sum_i\sum_j \exp\left(\frac{z_{it}\cdot z_{jt'}}{\|z_{it}\|_2\cdot\|z_{jt'}\|_2\cdot\tau}\right)}$$

(b) $\text{InfoNCE} = -\sum_{i,j} y_{itjt'} \log(\text{softmax}(\text{sim}(z_{it}, z_{jt'})/\tau))$

(c)
$$= -\sum_{i,j}\delta_{ij}\left[\frac{z_{it}\cdot z_{jt'}}{\|z_{it}\|_2\cdot\|z_{jt}\|_2\cdot\tau} - \log\left\{\sum_{i,j'}\exp\left(\frac{z_{it}\cdot z_{jt'}}{\|z_{it}\|_2\cdot\|z_{jt'}\|_2\cdot\tau}\right)\right\}\right]$$

$$= -\sum_{i=1}^{n}\left(\frac{z_{it}\cdot z_{it'}}{\|z_{it}\|_2\cdot\|z_{it'}\|_2\cdot\tau}\right) + n\log\left\{\sum_{i,j}\exp\left(\frac{z_{it}\cdot z_{jt'}}{\|z_{it}\|_2\cdot\|z_{jt'}\|_2\cdot\tau}\right)\right\}$$

$$= -\sum_{i=1}^{n}\frac{\text{sim}(z_{it}, z_{it'})}{\tau} + n\log\left[\sum_{i,j}\exp\left(\frac{\text{sim}(z_{it}, z_{jt'})}{\tau}\right)\right]$$

(d) First term implies that, for an original observation, z_i, loss is smaller if two different tranformation t and t' give similar mapping.

Second term shows that, for different original observations (z_i, z_j), loss is higher if two different transformation t and t' give similar mapping ~~(i.e. penelized~~

This second ~~trem~~ term prevents mode collapes.

h is used as repersentation because z is the result after projection, or demmension reduction.

~~3. (a) $K = Q = X + E$:~~ ~~$X = (x_{11}, x_{2}, \dots, x_n)$~~

~~$(K^TQ)_{i,j} = (X^T+E^T)(X+E)_{i,j} = (X^TX + E^TX + X^TE + E^TE)_{i,j}$~~

~~$= \sum_{i=1}^{n} \vec{x}_i \cdot \vec{x}_i +$~~ ~~$= \sum_{k=1}^{p} (X_{ki}X_{kj} + \sum_{k=1}^{p} E_{ki}X_{kj} + X$~~

~~$= \sum_{k=0}^{p-1} (X_{ki}X_{kj} + E_{ki}E_{kj} + X_{ki}E_{kj} + E_{ki}X_{kj})$~~

$= \sum_{k=0}^{p-1} \left(\sin\left(i\cdot\exp\left(-\frac{2k}{p}\log(10000\right.\right.\right.$

3. (c) (i) Addition : $K = X + E$.

$$K^T Q = (X^T + E^T)(X + E)$$

$$= X^T X + E^T X + E^T E + X^T E.$$

(ii) concatenation : $K = \begin{pmatrix} X \\ E \end{pmatrix}$.

$$K^T Q = (X^T, E^T)\begin{pmatrix} X \\ E \end{pmatrix}$$

$$= X^T X + E^T E.$$ No cross terms. $E^T X + X^T E$.

(d) $\cancel{f(X_i) = y_i = \begin{pmatrix} y_{i0} \\ y_{i1} \\ \vdots \\ y_{id} \\ 0 \\ 0 \end{pmatrix}}$ $\cancel{AX = \begin{pmatrix} Y \\ 0 \end{pmatrix}}$,

$\cancel{Ax_i = y_i}$ $\cancel{GE = \begin{pmatrix} F \\ 0 \end{pmatrix}}$. $GE = \begin{pmatrix} I_{f\times f} & 0 \\ 0 & 0 \end{pmatrix} = F$

$\cancel{E^T X = A =}$ $AX = \begin{pmatrix} I_{d\times d} & 0 \\ 0 & 0 \end{pmatrix} = Y$

$$E^T X = (G^{-1} G E)^T (A^{-1} A X) = F G^{-1} A^{-1} Y$$

$$H = G^{-1} A^{-1} = \begin{pmatrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{pmatrix}$$

$$HY = \begin{pmatrix} H_{00}{}_{f\times d} & H_{01} \\ H_{10} & H_{11} \end{pmatrix}\begin{pmatrix} I_{d\times d} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} H_{00} & 0 \\ H_{10} & 0 \end{pmatrix}$$

$$FHY = \begin{pmatrix} I_f & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} H_{00} \\ H_{10} \end{pmatrix} = \begin{pmatrix} H_{00}\,(f\times d) & 0 \\ 0 & 0 \end{pmatrix}$$

$$Y^T H^T F^T = \begin{pmatrix} H_{00}^T\,(d\times f) & 0 \\ 0 & 0 \end{pmatrix}.$$

$$E^T X + X^T E = \begin{pmatrix} H_{00}^T\,(d\times f) & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} H_{00}\,(f\times d) & 0 \\ 0 & 0 \end{pmatrix}$$

rank$(X) = d$

rank$(E) = f$.