8.b) Calculate gradient of SNE cost function with respect to y_i.

→ The SNE cost function is defined as:

$$C = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

gradient of C with respect to Y_i

chain rule- $$\frac{\delta}{\delta Y_i}\left(\log \frac{p_{j|i}}{q_{j|i}}\right) = \frac{1}{\frac{p_{j|i}}{q_{j|i}}} \cdot \frac{\delta}{\delta Y_i}\left(\frac{p_{j|i}}{q_{j|i}}\right)$$

$$\frac{\delta}{\delta Y_i}\left(\frac{p_{j|i}}{q_{j|i}}\right) = \frac{1}{q_{j|i}} \cdot \frac{\delta p_{j|i}}{\delta Y_i} - \frac{p_{j|i}}{q^2_{j|i}} \cdot \frac{\delta q_{j|i}}{\delta Y_i}$$

using gradient descent formula,

$$\frac{\delta}{\delta Y_i}\left(p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}\right) = 2 \cdot (p_{j|i} - q_{j|i}) \cdot (Y_i - Y_j)$$

summing over neighbors:

$$\frac{\delta C}{\delta Y_i} = \sum_j \frac{\delta}{\delta Y_i}\left(p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}\right)$$

$$\frac{\delta C}{\delta Y_i} = \sum_j 2 \cdot (p_{j|i} - q_{j|i}) \cdot (Y_i - Y_j)$$

Finally update Y_i using gradient descent update rule.

c) gradient descent of Symmetric SNE Cost function with respect to y_i.

For Symmetric SNE, there are further simplification to be made. Both p and q matrices are symmetric, so $p_{ij} = p_{ji}$ and $q_{ij} = q_{ji}$ leading to,

$$\frac{\delta C}{\delta y_i} = 2\sum_j (p_{ij} - q_{ij} + p_{ji} - q_{ji})(y_i - y_j)$$

$$= 2\sum_j (2p_{ij} - 2q_{ij})(y_i - y_j)$$

$$= 4\sum_j (p_{ij} - q_{ij})(y_i - y_j)$$

d) gradient descent of t-SNE.

$$\frac{\delta C}{dy_i} = 2\sum_j (k_{ij} + k_{ji})(y_i - y_j)$$

$$k_{ij} = \frac{1}{s}\left[\frac{\delta C}{\delta q_{ij}} - \sum \frac{\delta C}{\delta q_{kl}} q_{kl}\right] \frac{\delta w_{ij}}{\delta f_{ij}}$$

Insesting Kullback divergence

Both SNE and t-SNE use the Kullback-Leibles divergence, which as noted above, has following gradient:

$$\frac{\delta C}{\delta q_{ij}} = -\frac{p_{ij}}{q_{ij}}$$

k_{ij} therefore becomes:

$$k_{ij} = \frac{1}{s}\left[-\frac{p_{ij}}{q_{ij}} - \sum_{kl} \frac{-p_{kl}}{q_{kl}} q_{kl}\right] \frac{\delta w_{ij}}{\delta f_{ij}} = \frac{1}{s}\left[\frac{-p_{ij}}{q_{ij}} + \sum_{kl} p_{kl}\right] \frac{\delta w_{ij}}{\delta f_{ij}}$$

$$= \frac{1}{S}\left[-\frac{p_{ij}}{q_{ij}} + 1\right]\frac{\delta w_{ij}}{\delta f_{ij}}$$

At this point both SNE, and t-SNE output kernel (Gaussian and t-distribution respectively) have a desirable that has general form

$$\frac{\delta w_{ij}}{\delta f_{ij}} = -w_{ij}^{n}$$

where, $n=1$ in case of SNE, and $n=2$ in case of t-SNE

substituting that, we get.

$$k_{ij} = \frac{1}{S}\left[-\frac{p_{ij}}{q_{ij}} + 1\right]\frac{\delta w_{ij}}{\delta f_{ij}} = -\frac{w_{ij}^{n}}{S}\left[-\frac{p_{ij}}{q_{ij}} + 1\right] = -w_{ij}^{n-1} q_{ij}\left[\frac{p_{ij}}{q_{ij}}\right.$$

using the fact that $w_{ij}/S = q_{ij}$, now we can move $-q_{ij}$ inside the expression in parentheses to get:

$$k_{ij} = w_{ij}^{n-1}(p_{ij} - q_{ij})$$

for SNE $w_{ij}^{n-1} = 1$ because $n=1$ we get

$$\frac{\delta C}{\delta y_i} = 2\sum_j (p_{ij} - q_{ij} + p_{ji} - q_{ji})(Y_i - Y_j)$$

for t-SNE

$$\frac{\delta C}{\delta Y_i} = 2\sum_j \left[w_{ij}(p_{ij} - q_{ij}) + w_{ji}(p_{ji} - q_{ji})\right](Y_i - Y_j)$$

But t-distributed kernel also creates a symmetric matrix so we simplify

$$\frac{\delta C}{\delta Y_i} = 4\sum_j w_{ij}(p_{ij} - q_{ij})(Y_i - Y_j)$$

9.4

→ Consider the fitted values, that results from performing linear regression without an intercept. In this setting the i^{th} fitted value takes the form

$$\hat{y}_i = x_i \hat{\beta}$$

where,

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i'=1}^{n} x_{i'}^2}$$

now we want to express $\hat{y}_i$ as a linear comb^n of y_i's. we need to find coeff. $a_{i'}$ such that

$$\hat{y}_i = \sum_{i'=1}^{n} a_{i'} y_{i'}$$

we can rewrite the expression as;

$$\hat{y}_i = \frac{x_i}{\sum_{i'=1}^{n} x_{i'}^2} \left(\sum_{i=1}^{n} x_i y_i \right)$$

$$a_i = \frac{x_i}{\sum_{i'=1}^{n} x_{i'}^2}$$

Therefore, we can express $\hat{y}_i$ as a linear combination of y_i's.

$$\hat{y}_i = \sum_{i'=1}^{n} a_{i'} y_{i'}$$

the coeff $a_{i'}$ are given by

$$a_{i'} = \frac{x_{i'}}{\sum_{i'=1}^{n} x_{i'}^2}$$

This shows that fitted values from linear regression are indeed linear combination of response values $y_{i'}$.