1 Formatting

Input table T_{in} is formatted such that each row corresponds to one test. The first three columns in a row holds the hashes (h_1, h_2, h_3) for the audio samples in the test while the fourth column holds an integer $i \in [1, 4]$ denoting which of the samples was chosen as the "odd" one. The value 4 indicates that the test person could not tell:

h_1	h_2	h_3	i

We create a new, 4-column table, T_f, using all rows for which $i \neq 4$. In this table each test results in two rows, signifying the distances we want to compare. The two first columns always hold the non-odd samples while the third and fourth pairs the odd sample up against each of the two others. Renaming the samples $x_{1..3}$ we take the convention that x_3 is always the odd one, so each test gives us the following two rows:

x_1	x_2	x_1	x_3
x_1	x_2	x_2	x_3

For each row in T_{format} we obtain the distances d_1 and d_2, which denote the distances between the similar and dissimilar samples, respectively. Hence we want the algorithm to give us distances such that $d_1 < d_2$ whenever possible.

2 Cost

Having a distance function $d(...)_\theta$, we define the probability of seeing each test result as such:

$$p(? \mid d_1, d_2, \theta) = \Phi\Big(\frac{d_2 - d_1}{\sigma^2}\Big)$$

Hence, the larger d_2 is compared to d_1 the greater the chance of observing the test result. The combined probability is then

$$\prod(p(? \mid d_1, d_2, \theta))$$

for which we take the negative natural logarithm:

$$\begin{aligned} cost &= -\log\big(\prod(p(? \mid d_1, d_2, \theta))\big) \\ &= -\sum \log(p(? \mid d_1, d_2, \theta)) \end{aligned}$$

We the want to minimize *cost*, which lies in the interval $[0; \infty]$.

3 Gradient

Next we consider the gradient for our cost function,

$$\begin{aligned}
\frac{\partial cost}{\partial \mathbf{K}} &= \frac{\partial}{\partial \mathbf{K}} - \sum \log(p(? \mid d_1, d_2, \theta)) \\
&= \frac{\partial}{\partial \mathbf{K}} - \sum \log\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right) \\
&= -\frac{\partial}{\partial \mathbf{K}} \sum \log\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right) \\
&= -\sum testgrad \\
\text{where } testgrad &= \frac{\partial}{\partial \mathbf{K}} \log\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right)
\end{aligned}$$

Further derivation of *testgrad* yields:

$$\begin{aligned}
testgrad &= \frac{\partial}{\partial \mathbf{K}} \log\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right) \\
&= \log'\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right) \cdot \frac{\partial}{\partial \mathbf{K}}\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right) \\
&= \frac{1}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)} \cdot \frac{\partial}{\partial \mathbf{K}}\left(\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\right) \\
&= \frac{1}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)} \cdot \Phi'\left(\frac{d_2 - d_1}{\sigma^2}\right) \cdot \frac{\partial}{\partial \mathbf{K}} \frac{d_2 - d_1}{\sigma^2} \\
&= \frac{1}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)} \cdot \phi\left(\frac{d_2 - d_1}{\sigma^2}\right) \cdot \frac{\partial}{\partial \mathbf{K}} \frac{d_2 - d_1}{\sigma^2} \\
&= \frac{1}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)} \cdot \phi\left(\frac{d_2 - d_1}{\sigma^2}\right) \frac{1}{\sigma^2} \frac{\partial}{\partial \mathbf{K}} (d_2 - d_1) \\
&= \frac{1}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)} \cdot \phi\left(\frac{d_2 - d_1}{\sigma^2}\right) \frac{1}{\sigma^2} \left(\frac{\partial d_2}{\partial \mathbf{K}} - \frac{\partial d_1}{\partial \mathbf{K}}\right)
\end{aligned}$$

We separately consider the derivate of the distance function for vectors x_m and x_n:

$$\begin{aligned}\frac{\partial d}{\partial \mathbf{K}} &= \frac{\partial}{\partial \mathbf{K}}(x_m - x_n)^{\mathrm{T}}\mathbf{K}(x_m - x_n)\\ &= \frac{\partial}{\partial \mathbf{K}}(x_m - x_n)^{\mathrm{T}}\mathbf{L}\mathbf{L}^{\mathrm{T}}(x_m - x_n)\\ &= (x_m - x_n)^{\mathrm{T}}\frac{\partial \mathbf{L}\mathbf{L}^{\mathrm{T}}}{\partial \mathbf{K}}(x_m - x_n)\end{aligned}$$

We then have two cases.

If $\mathbf{K}$ is diagonal:

$$\frac{\partial \mathbf{L}\mathbf{L}^{\mathrm{T}}}{\partial \mathbf{K}} = \mathbf{I}$$

If $\mathbf{K}$ is dense:

$$\begin{aligned}\frac{\partial \mathbf{L}\mathbf{L}^{\mathrm{T}}}{\partial \mathbf{K}} &= \frac{\partial \mathbf{L}}{\partial \mathbf{L}}\mathbf{L}^{\mathrm{T}} + \mathbf{L}\frac{\partial \mathbf{L}^{\mathrm{T}}}{\partial \mathbf{L}}\\ &= \mathbf{L}^{\mathrm{T}} + \mathbf{L}\end{aligned}$$

For x_1, x_2, and x_3 we thus have the following combined gradient:

If $\mathbf{K}$ is diagonal:

$$\frac{\partial cost}{\partial \mathbf{K}} = -\sum \frac{\phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\Big((x_3 - x_1)^{\mathrm{T}}(x_3 - x_1) - (x_2 - x_1)^{\mathrm{T}}(x_2 - x_1)\Big)}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\sigma^2}$$

If $\mathbf{K} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$:

$$\frac{\partial cost}{\partial \mathbf{K}} = -\sum \frac{\phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\Big((x_3 - x_1)^{\mathrm{T}}(\mathbf{L}^{\mathrm{T}} + \mathbf{L})(x_3 - x_1) - (x_2 - x_1)^{\mathrm{T}}(\mathbf{L}^{\mathrm{T}} + \mathbf{L})(x_2 - x_1)\Big)}{\Phi\left(\frac{d_2 - d_1}{\sigma^2}\right)\sigma^2}$$