# Calculating the gradient of a cosine similarity loss function

With $v$, $w$ vectors, the cosine similarity is defined as

$$sim\left(x,y\right)\frac{v\cdot w}{\sqrt{v\cdot v*w\cdot w}}$$

Now with a matrix transformation $v = Mx$, $w = My$, we have a cosine similarity of

$$sim(x,y,M)=\frac{My\cdot Mx}{\sqrt{Mx\cdot Mx*My\cdot My}}$$

We want to calculate the derivatives wrt $M$. First we note

$$\frac{\partial v_k}{\partial M_{ij}}=\delta_{ik}x_j$$
$$\frac{\partial w_k}{\partial M_{ij}}=\delta_{ik}y_j$$

$$\frac{\partial\left(a\cdot z\right)}{\partial M_{ij}}=\sum_k\frac{\partial a_k}{\partial M_{ij}}z_k+a_k\frac{\partial z_k}{\partial M_{ij}}$$

Applying this, we have

$$\begin{aligned}
\frac{\partial\left(\frac{v\cdot w}{\sqrt{v\cdot v*w\cdot w}}\right)}{\partial M_{ij}} &= \frac{\frac{\partial(v\cdot w)}{\partial M_{ij}}}{\sqrt{v\cdot v*w\cdot w}}-\frac{v\cdot w}{2\left(v\cdot v*w\cdot w\right)^{\frac{3}{2}}}\frac{\partial\left(v\cdot v*w\cdot w\right)}{\partial M_{ij}}\\
&= \frac{\sum_k\delta_{ik}x_jw_k+v_k\delta_{ik}y_j}{\sqrt{v\cdot v*w\cdot w}}-\frac{v\cdot w}{2\left(v\cdot v*w\cdot w\right)^{\frac{3}{2}}}\left(\sum_k 2\delta_{ik}x_jv_k*w\cdot w+v\cdot v*\sum_k 2\delta_{ik}y_jw_k\right)\\
&= \frac{x_jw_i+v_iy_j}{\sqrt{v\cdot v*w\cdot w}}-\frac{v\cdot w}{\left(v\cdot v*w\cdot w\right)^{\frac{3}{2}}}\left(x_jv_i*w\cdot w+v\cdot v*y_jw_i\right)\\
&= \frac{x_jw_i+v_iy_j}{\left(v\cdot v*w\cdot w\right)^{3/2}}\left(v\cdot v*w\cdot w\right)-\frac{v\cdot w}{\left(v\cdot v*w\cdot w\right)^{\frac{3}{2}}}\left(x_jv_i*w\cdot w+v\cdot v*y_jw_i\right)\\
&= \frac{\left(x_jw_i+v_iy_j\right)v\cdot v*w\cdot w-v\cdot w\left(x_jv_i*w\cdot w+v\cdot v*y_jw_i\right)}{\left(v\cdot v*w\cdot w\right)^{3/2}}
\end{aligned}$$

This last formula is the one that is coded

# With Normalized vectors

If we normalize $x$ and $y$ before computation, we note this does not change the cosine similarity of the original vectors or the transformed vectors

## As a Hermitian Inner Product

We note that

$$\frac{My \cdot Mx}{\sqrt{Mx \cdot Mx * My \cdot My}} = \frac{y^T M^T Mx}{\sqrt{x^T M^T Mx * y^T M^T My}}$$

We note that $M$ only shows up in the form $M^T M$. So we actually don't need to treat $M$ as a transformation, but a normal inner product $A = M^T M$. We note that $M^T M$ will be symmetric for any $M$. Thus instead of optimizing over $M$ as an arbitrary matrix, we optimize over $A$ as a symmetric matrix, reducing the number of parameters by half. Thus the similarity measure is

$$sim\left(x, y, A\right) = \frac{y^T Ax}{\sqrt{x^T Ax * y^T Ay}}$$

To optimize over $A$, we need to calculate the derivative of $sim$ wrt the components of $A$. First we note

$$\frac{\partial a^T Ab}{\partial A_{ij}} = a_i b_j$$

$$\begin{aligned}
\frac{\partial sim}{\partial A_{ij}} &= \frac{y_i x_j}{\sqrt{x^T Ax * y^T Ay}} - \frac{y^T Ax}{2\left(x^T Ax * y^T Ay\right)^{\left(\frac{3}{2}\right)}}\left(x_i x_j y^T Ay + x^T Ax y_i y_j\right) \\
&= \frac{2y_i x_j \left(x^T Ax * y^T Ay\right) - y^T Ax \left(x_i x_j y^T Ay + x^T Ax y_i y_j\right)}{2\left(x^T Ax * y^T Ay\right)^{\left(\frac{3}{2}\right)}} \\
&= \frac{2y_i x_j \left(x^T Ax * y^T Ay\right) - y^T Ax \left(x_i x_j y^T Ay + x^T Ax y_i y_j\right)}{2\left(x^T Ax * y^T Ay\right)^{\left(\frac{3}{2}\right)}}
\end{aligned}$$

We can simplify this calculation by defining $n = x^T Ax$, $m = y^T Ay$, $l = x^T Ay$. Yielding

$$sim\left(x, y, A\right) = \frac{l}{\sqrt{nm}}$$

$$\frac{\partial sim}{\partial A_{ij}} = \frac{2y_i x_j nm - lmx_i x_j - lny_i y_j}{2\left(nm\right)^{\left(\frac{3}{2}\right)}}$$

If we consider the fact that the parameters $A_{ij}$ and $A_{ji}$ are tied together, we can calculate a total gradient for the off-diagonal elements of

$$\frac{dsim}{dA_{ij}} = \frac{2\left(y_i x_j + y_j x_i\right) nm - 2lmx_i x_j - 2lny_i y_j}{2\left(nm\right)^{\left(\frac{3}{2}\right)}} = \frac{nm\left(y_i x_j + y_j x_i\right) - lmx_i x_j - lny_i y_j}{\left(nm\right)^{\left(\frac{3}{2}\right)}}$$

And for the on-diagonal elements, we get

$$\frac{dsim}{dA_{ii}} = \frac{nmy_i x_i - \frac{1}{2}lmx_i^2 - \frac{1}{2}lny_i^2}{(nm)^{\left(\frac{3}{2}\right)}}$$