Partial Derivative of Gaussian Kernel Distance

Dakota Y. Hawkins

March 4, 2019

1 Identities

Let X be an $(N \times M)$ data matrix where N is the number of samples and M is the number of features. Then, x_i is an M length feature vector for sample i. Likewise, let $\vec{w}$ be an M element vector where each w_l is a feature weight describing the importance of feature l. Let y_{ij} represent the class identity function, such that $y_{ij} = 1$ if and only if $class(i) = class(j)$

$$D_{\vec{w}}(x_i, x_j) = \sum_{l=1}^{M} w_l^2 \left(x_{il} - x_{jl}\right)^2 \tag{1}$$

$$\bar{d} = \frac{\sum_{k \neq i}^{N} D_{\vec{w}}(x_i, x_k)}{(N-1)} \tag{2}$$

$$\bar{D}_{\vec{w}}(x_i, x_j, \bar{d}) = D_{\vec{w}}(x_i, x_j) - \bar{d} \tag{3}$$

$$K(\bar{D}_{\vec{w}}(x_i, x_j, \bar{d})) = \exp\left(-\frac{D_{\vec{w}}(x_i, x_j, \bar{d})}{\sigma}\right) \tag{4}$$

$$p_{ij} = \frac{K(\bar{D}_{\vec{w}}(x_i, x_j, \bar{d}))}{\sum_{k \neq i}^{N} K(\bar{D}_{\vec{w}}(x_i, x_k, \bar{d}))} \tag{5}$$

$$p_i = \sum_{j=1}^{N} y_{ij} p_{ij} \tag{6}$$

$$\bar{d}_l = \frac{\sum_{k \neq i}^{N} (x_{il} - x_{kl})^2}{(N-1)} \tag{7}$$

$$\xi(\vec{w}) = \sum_{i}^{N} \sum_{j \neq i}^{N} y_{ij} p_{ij} - \lambda \sum_{l}^{M} w_l^2 \tag{8}$$

- Equation 1 is the weighted euclidean distance between samples i and j.
- Equation 2 represents the average weighted distance between sample i and all other samples.

- Equation 3 is the weighted distance between samples i and j centered around the average distance from i
- Equation 4 is the kernel distance between samples.
- Equation 5 describes the probability of selecting sample j.
- Equation 6 is the probability of sample i being correctly classified.
- Equation 7 is the average distance between samples i and j in feature l.
- Equation 8 is the objective function we wish to maximize.

2 Partial of Kernel Distance

Taking the partial derivative of 4 with respect to w_l, for brevity, let $\bar{D}_{\vec{w}}(x_i, x_j, \bar{d}) = \bar{D}$ and begin by finding $\frac{\partial \bar{D}}{\partial w_l}$:

$$\begin{aligned}
\frac{\partial \bar{D}}{\partial w_l} &= \frac{\partial}{\partial w_l}\left(D_{\vec{w}}(x_i, x_j) - \bar{d}\right) \\
&= \frac{\partial}{\partial w_l} D_{\vec{w}}(x_i, x_j) - \frac{\partial}{\partial w_l}\bar{d} \\
&= 2w_l(x_{il} - x_{jl})^2 - \frac{2w_l}{(N-1)}\sum_{k\neq i}^{N}(x_{il} - x_{kl})^2 \\
\frac{\partial \bar{D}}{\partial w_l} &= 2w_l\left((x_{il} - x_{jl})^2 - \bar{d}_l\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial K(\bar{D})}{\partial w_l} &= \frac{\partial}{\partial w_l}\exp\left(-\frac{\bar{D}}{\sigma}\right) \\
&= \frac{\partial}{\partial \bar{D}}\exp\left(-\frac{\bar{D}}{\sigma}\right)\frac{\partial \bar{D}}{\partial w_l} \\
&= -\frac{1}{\sigma}\exp\left(-\frac{\bar{D}}{\sigma}\right)\frac{\partial \bar{D}}{\partial w_l} \\
\frac{\partial K(\bar{D})}{\partial w_l} &= -\frac{2w_l\left((x_{il} - x_{jl})^2 - \bar{d}_l\right)}{\sigma}K(\bar{D})
\end{aligned}$$

It immediately follows:

$$\frac{\partial}{\partial w_l}\sum_{k\neq i}^{N} K(\bar{D}_{\vec{w}}(x_i, x_k, \bar{d})) = \frac{-2w_l}{\sigma}\sum_{k\neq i}^{N}\left((x_{il} - x_{kl})^2 - \bar{d}_l\right)K(\bar{D}_{\vec{w}}(x_i, x_k, \bar{d}))$$

3 Partial of p_{ij}

Again for brevity, let $f_1 = K(\bar{D}_{\vec{w}}(x_i, x_j, \bar{d}))$ and $f_2 = \sum_{k \neq i}^{N} K(\bar{D}_{\vec{w}}(x_i, x_k, \bar{d}))$. Then:

$$\begin{aligned}
\frac{\partial p_{ij}}{\partial w_l} &= \frac{\partial}{\partial w_l} \left(\frac{K(\bar{D}_{\vec{w}}(x_i, x_j, \bar{d}))}{\sum_{k \neq i}^{N} K(\bar{D}_{\vec{w}}(x_i, x_k, \bar{d}))} \right) \\
&= \frac{\partial}{\partial w_l} \frac{f_1}{f_2} \\
\frac{\partial p_{ij}}{\partial w_l} &= \frac{\partial f_1}{\partial w_l} f_2^{-1} + f_1 \frac{\partial f_2^{-1}}{\partial w_l}
\end{aligned}$$

Plugging in the partial derivative for kernel distances found in section 2:

$$\begin{aligned}
\frac{\partial f_1}{\partial w_l} f_2^{-1} &= -\frac{2w_l \left((x_{il} - x_{jl})^2 - \bar{d}_l \right)}{\sigma} \frac{f_1}{f_2} \\
\frac{\partial f_1}{\partial w_l} f_2^{-1} &= -\frac{2w_l \left((x_{il} - x_{jl})^2 - \bar{d}_l \right)}{\sigma} p_{ij}
\end{aligned}$$

Plugging in the partial derivative for the sum of kernel distances also found in secion 2:

$$\begin{aligned}
f_1 \frac{\partial f_2}{\partial w_l} &= -\frac{f_1}{f_2^2} \frac{\partial}{\partial w_l} f_2 \\
&= -\frac{p_{ij}}{f_2} \frac{\partial}{\partial w_l} f_2 \\
&= \frac{2w_l p_{ij}}{\sigma f_2} \sum_{k \neq i}^{N} \left((x_{il} - x_{kl})^2 - \bar{d}_l \right) K(\bar{D}_{\vec{w}}(x_i, x_k, \bar{d}) \\
&= \frac{2w_l p_{ij}}{\sigma} \sum_{k \neq i}^{N} \left((x_{il} - x_{kl})^2 - \bar{d}_l \right) \frac{f_{1k}}{f_2} \\
f_1 \frac{\partial f_2}{\partial w_l} &= \frac{2w_l p_{ij}}{\sigma} \sum_{k \neq i}^{N} \left((x_{il} - x_{kl})^2 - \bar{d}_l \right) p_{ik}
\end{aligned}$$

Combining terms:

$$\begin{aligned}
\frac{\partial p_{ij}}{\partial w_l} &= \frac{\partial f_1}{\partial w_l} f_2^{-1} + f_1 \frac{\partial f_2^{-1}}{\partial w_l} \\
&= -\frac{2w_l \left((x_{il} - x_{jl})^2 - \bar{d}_l \right)}{\sigma} p_{ij} + \frac{2w_l p_{ij}}{\sigma} \sum_{k \neq i}^{N} \left((x_{il} - x_{kl})^2 - \bar{d}_l \right) p_{ik} \\
\frac{\partial p_{ij}}{\partial w_l} &= \frac{2w_l p_{ij}}{\sigma} \left(\sum_{k \neq i}^{N} \left[\left((x_{il} - x_{kl})^2 - \bar{d}_l \right) p_{ik} \right] - \left((x_{il} - x_{jl})^2 - \bar{d}_l \right) \right)
\end{aligned}$$

4 Partial of objective function

$$
\begin{aligned}
\xi(\vec{w}) &= \sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}p_{ij} - \lambda\sum_{l}^{M} w_l^2 \\
\frac{\partial\xi(\vec{w})}{\partial w_l} &= \frac{\partial}{\partial w_l}\left(\sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}p_{ij} - \lambda\sum_{l}^{M} w_l^2\right) \\
&= \sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}\frac{\partial}{\partial w_l}p_{ij} - \lambda\sum_{l}^{M}\frac{\partial}{\partial w_l}w_l^2 \\
&= \sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}\frac{\partial p_{ij}}{\partial w_l} - 2\lambda w_l \\
&= \sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}\left[\frac{2w_l p_{ij}}{\sigma}\left(\sum_{k\neq i}^{N}\left[\left((x_{il}-x_{kl})^2-\bar{d}_l\right)p_{ik}\right] - \left((x_{il}-x_{jl})^2-\bar{d}_l\right)\right)\right] - 2\lambda w_l \\
&= \frac{2w_l}{\sigma}\left[\sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}\left(p_{ij}\sum_{k\neq i}^{N}\left[\left((x_{il}-x_{kl})^2-\bar{d}_l\right)p_{ik}\right] - \left((x_{il}-x_{jl})^2-\bar{d}_l\right)\right)\right] - 2\lambda w_l \\
&= \frac{2w_l}{\sigma}\left[\sum_{i}^{N}\sum_{j\neq i}^{N} y_{ij}p_{ij}\sum_{k\neq i}^{N}\left[\left((x_{il}-x_{kl})^2-\bar{d}_l\right)p_{ik}\right] - \sum_{j\neq i}^{N} y_{ij}p_{ij}\left((x_{il}-x_{jl})^2-\bar{d}_l\right)\right] - 2\lambda w_l \\
&= \frac{2w_l}{\sigma}\left[\sum_{i}^{N} p_i\sum_{k\neq i}^{N}\left[\left((x_{il}-x_{kl})^2-\bar{d}_l\right)p_{ik}\right] - \sum_{j\neq i}^{N} y_{ij}p_{ij}\left((x_{il}-x_{jl})^2-\bar{d}_l\right)\right] - 2\lambda w_l \\
\frac{\partial\xi(\vec{w})}{\partial w_l} &= 2w_l\left(\frac{1}{\sigma}\left[\sum_{i}^{N} p_i\sum_{k\neq i}^{N}\left[\left((x_{il}-x_{kl})^2-\bar{d}_l\right)p_{ik}\right] - \sum_{j\neq i}^{N} y_{ij}p_{ij}\left((x_{il}-x_{jl})^2-\bar{d}_l\right)\right] - \lambda\right)
\end{aligned}
$$