Partial Derivatives for NCFS Optimization

Dakota Y. Hawkins

February 10, 2020

1 Identities

From [?], the probability a sample j is used as a reference point for sample i in a leave-on-out KNN prediction is:

$$p_{ij} = \begin{cases} \dfrac{K(D_w(x_i, x_j))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_k))}, & \text{i} \neq j \\ 0, & \text{i = j} \end{cases} \tag{1}$$

Where X is a $(N \times P)$ data matrix, where N is the number of samples and P is the number of features. x_i is the row feature vector for sample i where $x_i = \langle x_{i1}, x_{i2}, \ldots, x_{ip} \rangle$, $i \in \{1, \ldots, N\}$. Let $D_w(x_i, x_j)$ be some weighted distance function to measure differences between sample vectors, such that w_l is the weight prescribed to the lth feature, $l \in \{1, \ldots, P\}$ $w = \langle w_1, \ldots, w_P \rangle$. In this document it is assumed that **all weights are squared** to ensure non-negative properties. Examples are listed in the next section Finally, let $K(D_w(x_i, w_j))$ be the kernel function from [?] where

$$K(D) = \exp\left(-\frac{D}{\sigma}\right) \tag{2}$$

for some real number σ.

In order to account for class balance, we modify the objective function in [?] from

$$\xi(w) = \sum_{i=1}^{N} \sum_{j=1}^{N} y_{ij} \cdot p_{ij} - \lambda \sum_{l=1}^{D} w_l^2 \tag{3}$$

to

$$\zeta(w) = \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} p_{ij} - \lambda \sum_{l=1}^{D} w_l^2 \tag{4}$$

where, given $c = \langle c_1, \cdots, c_N \rangle$, is a vector of class assignments for each sample i in X.

$$y_{ij} = \begin{cases} 1, & \text{c}_i = c_j \\ 0, & \text{c}_i \neq c_j \end{cases} \tag{5}$$

and C_i is the total number of samples with label C_i.

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

Finally, let p_i be the probability that sample i is is correctly assigned it's given class during KNN selection:

$$p_i = \sum_{j=1}^{N} y_{ij} p_{ij} \tag{7}$$

2 Distances

Below is a list of supported distances.

1. Manhattan, City Block, L1

$$D_w(x_i, x_j) = \sum_{l=1}^{P} w_l^2 \left| x_{il} - x_{jl} \right| \tag{8}$$

2. Euclidean, L2

$$D_w(x_i, x_j) = \sqrt{\sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2} \tag{9}$$

3. Squared Euclidean, Squared L2

$$D_w(x_i, x_j) = \sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2 \tag{10}$$

4. Φ_s

$$D_w(x_i, x_j) = \frac{Var_w(x_i - x_j)}{Var_w(x_i + x_j)} \tag{11}$$

Where

$$Var_w(x_i) = \frac{\sum_{l=1}^{P} w^2 (x_{il} - \mu_{x_i}^w)^2}{V_1 - \frac{V_2}{V_1}} \tag{12}$$

$$V_1 = \sum_{l=1}^{P} w_l^2 \tag{13}$$

$$V_2 = \sum_{l=1}^{P} w_l^4 \tag{14}$$

$$\mu_{x_i}^w = \frac{\sum_{l=1}^{P} x_{il} \cdot w_l^2}{V_1} \tag{15}$$

3 Partial Derivative of Objective Function

In this section we find the partial derivative of $\zeta(w)$ with respect to w_l for an arbitrary distance function.

$$\begin{aligned}
\zeta(w) = \zeta(w) &= \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} p_{ij} - \lambda \sum_{l=1}^{D} w_l^2 \\
\frac{\partial \zeta(w)}{\partial w_l} &= \frac{\partial}{\partial w_l} \left(\sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} p_{ij} - \lambda \sum_{l=1}^{D} w_l^2 \right) \\
&= \frac{\partial}{\partial w_l} \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} p_{ij} - \frac{\partial}{\partial w_l} \lambda \sum_{l=1}^{D} w_l^2 \\
&= \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} \frac{\partial p_{ij}}{\partial w_l} - \lambda \sum_{l=1}^{D} \frac{\partial w_l^2}{\partial w_l} \\
\frac{\partial \zeta(w)}{\partial w_l} &= \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} \frac{\partial p_{ij}}{\partial w_l} - 2 w_l \lambda
\end{aligned}$$

From (**??**), $p_{ij} = 0$ when $i = j$. It follows $\frac{\partial \zeta(w)}{\partial w_l} = 0$ when $i = j$. Before solving for $\frac{\partial \zeta(w)}{\partial w_l}$ when $i \neq j$, we first solve for the partial derivatives of $K\left(D_w(x_i, x_j)\right)$ for an arbitrary weighted distance function.

$$\begin{aligned}
K\left(D_w(x_i, x_j)\right) &= e^{\frac{-D_w(x_i, x_j)}{\sigma}} \\
\frac{\partial K\left(D_w(x_i, x_j)\right)}{\partial w_l} &= \frac{\partial}{\partial w_l} e^{\frac{-D_w(x_i, x_j)}{\sigma}} \\
&= \frac{-1}{\sigma} \frac{\partial D_w(x_i, x_j)}{\partial w_l} e^{\frac{-D_w(x_i, x_j)}{\sigma}} \\
\frac{\partial K\left(D_w(x_i, x_j)\right)}{\partial w_l} &= \frac{-1}{\sigma} \frac{\partial D_w(x_i, x_j)}{\partial w_l} K(D_w(x_i, x_j))
\end{aligned}$$

Extending above it is easy to see that:

$$\frac{\partial}{\partial w_l} \sum_{k=1}^{N} K(D_w(x_i, x_k)) = \sum_{k=1}^{N} \frac{-1}{\sigma} \frac{\partial D_w(x_i, x_k)}{\partial w_l} K(D_w(x_i, x_k))$$

Further, letting $f_1 = K\left(D_w(x_i, x_j)\right)$ and $f_2 = \sum_{k=1}^{N} K(D_w(x_i, x_k))$:

$$\frac{K(D_w(x_i,x_j))}{\sum_{k=1}^{N} K(D_w(x_i,x_k))} = \frac{f_1}{f_2}$$

$$= f_1(f_2)^{-1}$$

$$\frac{\partial}{\partial w_l}\frac{K(D_w(x_i,x_j))}{\sum_{k=1}^{N} K(D_w(x_i,x_k))} = \frac{\partial}{\partial w_l}\left(f_1(f_2)^{-1}\right)$$

$$= f_1\frac{\partial f_2^{-1}}{\partial w_l} + \frac{\partial f_1}{\partial w_l} f_2^- 1$$

Solving for in for $f_1 \frac{\partial}{\partial w_l} f_2^{-1}$

$$f_1\frac{\partial}{\partial w_l} f_2^{-1} = -K(D_w(x_i,x_j)) \cdot \frac{\sum_{k=1}^{N} \frac{\partial}{\partial w_l} K(D_w(x_i,x_k))}{(\sum_{k=1}^{N} K(D_w(x_i,x_k))^2}$$

Substituting in equation **??**:

$$f1\frac{\partial f_2^{-1}}{\partial w_l} = -p_{ij}\frac{\sum_{k=1}^{N} \frac{\partial}{\partial w_l} K(D_w(x_i,x_k))}{\sum_{k=1}^{N} K(D_w(x_i,x_j))}$$

Similarly solving for $\frac{\partial f_1}{\partial w_l} f_2^{-1}$:

$$\frac{\partial f_1}{\partial w_l} f_2^{-1} = \frac{\frac{\partial}{\partial w_l} K(D_w(x_i,x_j))}{\sum_{k=1}^{N} K(D_w(x_i,x_j))}$$

Now solving for $\frac{\partial}{\partial w_l} p_{ij}$ when $i \neq j$:

$$
\begin{aligned}
p_{ij} &= \frac{K(D_w(x_i, x_j))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_k))} \\
\frac{\partial p_{ij}}{\partial w_l} &= \frac{\partial}{\partial w_l}\left(\frac{K(D_w(x_i, x_j))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_k))}\right) \\
&= \frac{\partial}{\partial w_l}(f_1 f_2^{-1}) \\
&= f_1 \frac{\partial f_2^{-1}}{\partial w_l} + \frac{\partial f_1}{\partial w_l} f_2^- 1 \\
&= -p_{ij} \frac{\sum\limits_{k=1}^{N} \frac{\partial}{\partial w_l} K(D_w(x_i, x_k))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_j))} + \frac{\frac{\partial}{\partial w_l} K(D_w(x_i, x_j))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_j))} \\
&= p_{ij} \frac{\sum\limits_{k=1}^{N} \frac{\partial D_w(x_i,x_j)}{\partial w_l} K(D_w(x_i, x_k))}{\sigma \sum\limits_{k=1}^{N} K(D_w(x_i, x_k))} - \frac{\frac{\partial D_w(x_i,x_j)}{\partial w_l} K(D_w(x_i, x_j))}{\sigma \sum\limits_{k=1}^{N} K(D_w(x_i, x_k)} \\
&= p_{ij} \frac{\sum\limits_{k=1}^{N} \frac{\partial D_w(x_i,x_j)}{\partial w_l} K(D_w(x_i, x_k))}{\sigma \sum\limits_{k=1}^{N} K(D_w(x_i, x_k))} - \frac{\partial D_w(x_i, x_j)}{\partial w_l} \frac{p_{ij}}{\sigma} \\
&= \frac{p_{ij}}{\sigma}\left(\frac{\sum\limits_{k=1}^{N} \frac{\partial D_w(x_i,x_j)}{\partial w_l} K(D_w(x_i, x_k))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_k))} - \frac{\partial D_w(x_i, x_j)}{\partial w_l}\right) \\
&= \frac{p_{ij}}{\sigma}\left(\sum_{k=1}^{N} \frac{\partial D_w(x_i, x_j)}{\partial w_l} \frac{K(D_w(x_i, x_k))}{\sum\limits_{k=1}^{N} K(D_w(x_i, x_k))} - \frac{\partial D_w(x_i, x_j)}{\partial w_l}\right) \\
\frac{\partial p_{ij}}{\partial w_l} &= \frac{p_{ij}}{\sigma}\left(\sum_{k=1}^{N} \frac{\partial D_w(x_i, x_j)}{\partial w_l} p_{ik} - \frac{\partial D_w(x_i, x_j)}{\partial w_l}\right)
\end{aligned}
$$

Continuing to solve for $\frac{\partial \zeta(w)}{\partial w_l}$:

$$
\begin{aligned}
\frac{\partial \zeta(w)}{\partial w_l} &= \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} \frac{\partial p_{ij}}{\partial w_l} - 2w_l\lambda \\
&= \sum_{i=1}^{N} \frac{1}{C_i} \sum_{j=1}^{N} y_{ij} \left[\frac{p_{ij}}{\sigma} \left(\sum_{k=1}^{N} \frac{\partial D_w(x_i,x_j)}{\partial w_l} p_{ik} - \frac{\partial D_w(x_i,x_j)}{\partial w_l} \right) \right] - 2w_l\lambda \\
&= \frac{1}{\sigma} \sum_{i=1}^{N} \frac{1}{C_i} \left[\sum_{j=1}^{N} y_{ij} p_{ij} \sum_{k=1}^{N} \frac{\partial D_w(x_i,x_j)}{\partial w_l} p_{ik} - \sum_{j=1}^{N} y_{ij} p_{ij} \frac{\partial D_w(x_i,x_j)}{\partial w_l} \right] - 2w_l\lambda \\
\frac{\partial \zeta(w)}{\partial w_l} &= \frac{1}{\sigma} \sum_{i=1}^{N} \frac{1}{C_i} \left[p_i \sum_{k=1}^{N} \frac{\partial D_w(x_i,x_j)}{\partial w_l} p_{ik} - \sum_{j=1}^{N} y_{ij} p_{ij} \frac{\partial D_w(x_i,x_j)}{\partial w_l} \right] - 2w_l\lambda
\end{aligned}
$$

4 Partial Derivative of Distance Metrics

4.1 L1 Distance

$$
\begin{aligned}
D_w(x_i,x_j) &= \sum_{l=1}^{P} w_l^2 \left| x_{il} - x_{jl} \right| \\
\frac{\partial D_w(x_i,x_j)}{\partial w_l} &= \sum_{l=1}^{P} \frac{\partial}{\partial w_l} w_l^2 \left| x_{il} - x_{jl} \right| \\
\frac{\partial D_w(x_i,x_j)}{\partial w_l} &= w_l^2 \left| x_{il} - x_{jl} \right|
\end{aligned}
$$

4.2 L2 Distance

$$
\begin{aligned}
D_w(x_i,x_j) &= \sqrt{\sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2} \\
\frac{\partial D_w(x_i,x_j)}{\partial w_l} &= \frac{\partial}{\partial w_l} \sqrt{\sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2} \\
&= \frac{\partial}{\partial w_l} \left(\sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{2\sqrt{\sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2)}} \frac{\partial}{\partial w_l} \sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl} \right)^2 \\
\frac{\partial D_w(x_i,x_j)}{\partial w_l} &= \frac{w_l \left(x_{il} - x_{jl} \right)^2}{D_w(x_i,x_j)}
\end{aligned}
$$

4.3 Squared L2 Distance

$$
\begin{aligned}
D_w(x_i, x_j) &= \sum_{l=1}^{P} w_l^2 \left(x_{il} - x_{jl}\right)^2 \\
\frac{\partial D_w(x_i, x_j)}{\partial w_l} &= \sum_{l=1}^{P} \frac{\partial}{\partial w_l} w_l^2 \left(x_{il} - x_{jl}\right)^2 \\
\frac{\partial D_w(x_i, x_j)}{\partial w_l} &= 2w_l \left(x_{il} - x_{jl}\right)^2
\end{aligned}
$$