Supplementary Material to the Paper: Pairwise Learning to Rank by Neural Networks Revisited: Reconstruction, Theoretical Analysis and Practical Performance

1 The NDCG Metric

In the field of learning to rank, a commonly used measure for the performance of a model is the normalized discounted cumulative gain of top-k documents retrieved (NDCG@k). This metric is based on the discounted cumulative gain of top-k documents (DCG@k):

$$\mathrm{DCG@}k = \sum_{i=1}^{k} \frac{2^{r(d_i)} - 1}{log_2(i+1)},$$

where $d_1, d_2, ..., d_n$ is the list of documents sorted by the model with respect to a single query and $r(d_i)$ is the relevance label of document d_i. The NDCG@k can be computed by dividing the DCG@k by the ideal (maximum) discounted cumulative gain of top-k documents retrieved (IDCG@k), i.e. the DCG@k for a perfectly sorted list of documents is defined as NDCG@$k = \frac{\mathrm{DCG@}k}{\mathrm{IDCG@}k}$.

2 The MAP Metric

A frequently used alternative to the NDCG is the mean average precision (MAP). For this metric the precision for the top-k documents of query q is introduced. Since this only makes sense for binary classes, a multiclass system has to be binarized such that $r(d_i) = 1$ indicates a relevant, and $r(d_i) = 0$ indicates an irrelevant document:

$$\mathrm{P}_q\mathrm{@}k = \frac{1}{k}\sum_{i=1}^{k} r(d_i)$$

Now one needs to calculate the average precision over all *relevant* documents.

$$\mathrm{AvgP}_q = \frac{1}{n \cdot P\mathrm{@}n}\sum_{k=1}^{n} r(d_k)\mathrm{P}_q\mathrm{@}k$$

where n is the number of documents in the query q. Finally, the MAP is given by the mean over all queries:

$$\mathrm{MAP} = \frac{1}{Q}\sum_{q=1}^{Q} \mathrm{AvgP}_q$$

where Q is the number of queries in a data set.

3 The Proof for Theorem 4

Theorem 1. *Let $\succeq$ be reflexive, antisymmetric, and transitive and $\mathcal{F} \subset \mathbb{R}^n$ be convex and open. For every $x \in \mathcal{F}$ define*

$$\mathcal{P}_x := \{y \in \mathcal{F} | x \not\succeq y\}, \quad \mathcal{N}_x := \{y \in \mathcal{F} | y \not\succeq x\},$$
$$\partial_x := \{y \in \mathcal{F} | x \succeq y \wedge y \succeq x\}$$

Furthermore, let $(\mathcal{F}/\sim, d)$ be a metric space, where $x \sim y \Leftrightarrow y \in \partial_x$ and

$$d(\partial_x, \partial_y) = \inf_{x' \in \partial_x, y' \in \partial_y} \|x' - y'\|$$

Then, the DirectRanker can train $\succeq$ if $\mathcal{P}_x$ and $\mathcal{N}_x$ are open for all $x \in \mathcal{F}$.

Proof. First, note that $\partial_x = \mathcal{F} \backslash (\mathcal{P}_x \cup \mathcal{N}_x)$ and that $\mathcal{P}_x \cap \mathcal{N}_x = \emptyset$ because of the antisymmetry of $\succeq$, dividing $\mathcal{F}$ into three distinct subsets $\mathcal{P}_x, \mathcal{N}_x, \partial_x$. The relation $x \sim y \Leftrightarrow y \in \partial_x$ defines an equivalence relation with equivalence classes $[x] = \partial_x$.
According to the universal approximation theorem it is sufficient to show that a continuous function $g : \mathcal{F} \to \mathbb{R}$ exists such that $\forall x, y \in \mathcal{F} : x \succeq y \Leftrightarrow g(x) \geq g(y)$. Again this will be done by explicit construction:
Now, define $g : \mathcal{F} \to \mathbb{R}$ by first choosing an arbitrary $x_0 \in \mathcal{F}$. For all points, set

$$g(x) = \begin{cases} d(\partial_{x_0}, \partial_x) & \text{if } x \in \mathcal{P}_{x_0} \\ -d(\partial_{x_0}, \partial_x) & \text{if } x \in \mathcal{N}_{x_0} \\ 0 & \text{if } x \in \partial_{x_0} \end{cases} \quad \forall x \in \mathcal{F}$$

First, let us show, that g is continuous:
Consider $d_{\partial_{x_0}} : (\mathcal{F}/\sim) \to \mathbb{R}$, $\partial_x \mapsto d(\partial_{x_0}, \partial_x)$. Since d as a metric is continuous, we know that $d_{\partial_{x_0}}$ is, too. Thus, for every $\partial_x \in \mathcal{F}/\sim$ and every $\varepsilon > 0$ there is a $\delta > 0$ such that for every ∂_y with $d(\partial_x, \partial_y) < \delta$ $|d_{\partial_{x_0}}(\partial_x) - d_{\partial_{x_0}}(\partial_y)| < \varepsilon$. Therefore, for a given $\varepsilon > 0$ and $x \in \mathcal{F}$ we can always choose δ such that for every $y \in \mathcal{F}$ with $\|x - y\| < \delta$ the above holds for ∂_x and ∂_y. Then, $| |g(x)| - |g(y)| | = |d_{\partial_{x_0}}(\partial_x) - d_{\partial_{x_0}}(\partial_y)| < \varepsilon$. Thus, $|g|$ is continuous, which means that g is continuous on $\mathcal{P}_{x_0} \cup \partial_{x_0} = \mathcal{F} \backslash \mathcal{N}_{x_0}$ and $\mathcal{N}_{x_0} \cup \partial_{x_0} = \mathcal{F} \backslash \mathcal{P}_{x_0}$. Since these two sets are closed in $\mathcal{F}$, this implies that g is continuous on their union $\mathcal{F} = \mathcal{P}_{x_0} \cup \partial_{x_0} \cup \mathcal{N}_{x_0}$.
Finally, we need to show now that $r(x, y) = g(x) - g(y)$ represents $\succeq$, i.e. $x \succeq y \Leftrightarrow g(x) \geq g(y)$:
For the case $x \succeq x_0 \succeq y$ the equivalence is obvious from the definition of g. For the remaining cases concerning $x \succeq y \Rightarrow g(x) \geq g(y)$, suppose there are elements $x, y \in \mathcal{P}_{x_0}$ with $x \succeq y$ and $g(x) < g(y)$. We can then choose a continuous curve $\gamma : [0, 1] \to \mathcal{F}$ with $\gamma(0) = x_0 \in \mathcal{N}_y$ and $\gamma(1) = x \in \mathcal{P}_y \cup \partial_y$. Choose $t_0 := \sup\{t \in [0, 1] | \gamma(t) \in \mathcal{N}_y\}$. Then, for every neighborhood U of t_0, there are $t', t'' \in U$, such that $\gamma(t') \in \mathcal{N}_y$, $\gamma(t'') \notin \mathcal{N}_y$. Therefore, $\gamma(t_0)$ is a boundary point of N_y and therefore no element of N_y. Also, since $\mathcal{P}_y \cap \mathcal{N}_y = \emptyset$, $\gamma(t_0)$ is no inner point of P_y and therefore no element of P_y. Thus $\gamma(t_0) =: \tilde{y} \in \mathcal{F} \backslash (\mathcal{P}_y \cup \mathcal{N}_y) = \partial_y$. If $y \succeq x$, t_0 may equal 1.
Since this holds for any continuous curve γ with the given boundary conditions, we can choose $\gamma(t) = tx + (1 - t)x_0$. The length of $\gamma|_{[0,t_0]}$ is then given by

$$\|\tilde{y} - x_0\| = L(\gamma|_{[0,t_0]}) = \int_0^{t_0} \|\dot{\gamma}(t)\| dt \leq \int_0^1 \|\dot{\gamma}(t)\| dt = \|x - x_0\|$$

The same argument holds when replacing x and x_0 by arbitrary $x' \in \partial_x$ and $x'_0 \in \partial_{x_0}$. Therefore, $\forall x' \in \partial_x, x'_0 \in \partial_{x_0} \exists \tilde{y}_{x',x'_0} \in \partial_y : \|y_{x',x'_0} - x'_0\| \leq \|x' - x'_0\|$, and therefore

$$g(y) = d(\partial_y, \partial_{x_0}) = \inf_{y' \in \partial_y, x'_0 \in \partial_{x_0}} \|y' - x'_0\| \leq \inf_{x' \in \partial_x, x'_0 \in \partial_{x_0}} \|x' - x'_0\| = d(\partial_x, \partial_{x_0}) = g(x)$$

which contradicts the assumption that $g(x) < g(y)$.

Following a similar chain of reasoning, the case $g(x) < g(y)$ for $x, y \in \mathcal{N}_{x_0}$ leads to contradictions.
To see that $g(x) \geq g(y) \Rightarrow x \succeq y$, suppose $\exists x, y \in \mathcal{F}, g(x) \geq g(y), x \not\succeq y$. Because of the antisymmetry of $\succeq$, this implies $y \succeq x$, which (as shown above) leads to $g(y) \geq g(x)$. This however, is only possible if $g(x) = g(y)$.
Analogous to the above, for every curve $\gamma(t) = tx' + (1-t)x_0'$ for $t \in [0,1]$ with $x' \in \partial_x$, $x_0' \in \partial_{x_0}$, there exists a $\tilde{y}_{x',x_0'} \in \partial_y$ and $t_0 \in]0,1[$ such that $\gamma(t_0) = \tilde{y}_{x',x_0'}$. We also know that

$$\|x' - x_0'\| = \int_0^1 \|\dot{\gamma}(t)\| dt = \int_0^{t_0} \|\dot{\gamma}(t)\| dt + \int_{t_0}^1 \|\dot{\gamma}(t)\| dt = \|\tilde{y}_{x',x_0'} - x_0'\| + \|x' - \tilde{y}_{x',x_0'}\|$$

However, $g(x) = g(y) \Leftrightarrow \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - x_0'\| = \inf_{y' \in \partial_y, x_0' \in \partial_{x_0}} \|y' - x_0'\|$ implies

$$\inf_{y' \in \partial_y, x_0' \in \partial_{x_0}} \|y' - x_0'\| \leq \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|\tilde{y}_{x',x_0'} - x_0'\| \leq \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - x_0'\| \leq \inf_{y' \in \partial_y, x_0' \in \partial_{x_0}} \|y' - x_0'\|$$

This means, that $\inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|\tilde{y}_{x',x_0'} - x_0'\| = \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - x_0'\|$, which implies

$$\begin{aligned}
\inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - x_0'\| &= \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} (\|\tilde{y}_{x',x_0'} - x_0'\| + \|x' - \tilde{y}_{x',x_0'}\|) \\
&\geq \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|\tilde{y}_{x',x_0'} - x_0'\| + \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - \tilde{y}_{x',x_0'}\| \\
&\geq \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - x_0'\| + \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - \tilde{y}_{x',x_0'}\|
\end{aligned}$$

$$\begin{aligned}
\Rightarrow d(\partial_x, \partial_y) = \inf_{x' \in \partial_x, y' \in \partial_y} \|x' - y'\| &\leq \inf_{x' \in \partial_x, x_0' \in \partial_{x_0}} \|x' - \tilde{y}_{x',x_0'}\| \leq 0 \\
\Rightarrow d(\partial_x, \partial_y) &= 0
\end{aligned}$$

Which contradict the assumption $x \not\succeq y$ that implies $\partial_x \neq \partial_y$. $\square$