A Proofs of Theorems

Theorem 1. *Given a multi-view dataset $\mathcal{D}$, and let $\mathcal{S} \subset \mathcal{D}$ be a set of ψ sampled points. For each point $\mathbf{s}^v \in \mathcal{S}^v$, let $\theta(\mathbf{s}^v)$ denote the neighborhood generated by $\mathbf{s}^v$, and $\mathcal{R}(\mathbf{s}^v)$ denote the set of all normal instances from $\mathcal{D}$ that within $\theta(\mathbf{s}^v)$ in view v. Then, for any two views v_1 and v_2 of $\mathcal{D}$, $\mathcal{R}(\mathbf{s}^{v_1})$ and $\mathcal{R}(\mathbf{s}^{v_2})$ contain essentially the same instances, i.e.,*

$$\mathbb{E}[|\mathcal{R}(\mathbf{s}^{v_1})|] \simeq \mathbb{E}[|\mathcal{R}(\mathbf{s}^{v_2})|]$$

where $\mathbb{E}[|\cdot|]$ is the expected number of the set.

Proof. Since normal instances in multi-view data have largely similar neighborhood structures, for any normal instance $\mathbf{x} \in \mathcal{D}$, it follows:

$$P(\mathbf{x}^{v_1} \in \theta[\mathbf{s}^{v_1}]) \simeq P(\mathbf{x}^{v_2} \in \theta[\mathbf{s}^{v_2}]).$$

Let $\mathcal{N} \subset \mathcal{D}$ denote the set of normal instances in $\mathcal{D}$. Then,

$$\begin{aligned}\mathbb{E}[|\mathcal{R}(\mathbf{s}^{v_1})|] &= \sum_{\mathbf{x} \in \mathcal{N}} P(\mathbf{x}^{v_1} \in \theta[\mathbf{s}^{v_1}]) \\ &\simeq \sum_{\mathbf{x} \in \mathcal{N}} P(\mathbf{x}^{v_2} \in \theta[\mathbf{s}^{v_2}]) \\ &= \mathbb{E}[|\mathcal{R}(\mathbf{s}^{v_2})|]\end{aligned}$$

This result proves the theorem. □

Theorem 2. *Given two datasets $\mathcal{D}$ and $\mathcal{D}'$ with same number of points observed in view of v, where each point in $\mathcal{D}$ and $\mathcal{D}'$ belongs to a subspace $\mathcal{X} \subseteq \mathbb{R}^d$ and is drawn from probability distributions $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{D}'}$ defined on $\mathbb{R}^d$, respectively. Both $\mathcal{P}_{\mathcal{D}}$ and $\mathcal{P}_{\mathcal{D}'}$ are strictly positive on $\mathcal{X}$. Let $\mathcal{E} \subset \mathcal{X}$ be a region such that for all $\mathbf{x} \in \mathcal{E}$, $\mathcal{P}_{\mathcal{D}}(\mathbf{x}) < \mathcal{P}_{\mathcal{D}'}(\mathbf{x})$, i.e., $\mathcal{D}$ is sparser than $\mathcal{D}'$ in $\mathcal{E}$. Given two randomly sampled sets $\mathcal{S} \subset \mathcal{D}$ and $\mathcal{S}' \subset \mathcal{D}'$, where $|\mathcal{S}| = |\mathcal{S}'| = \psi$. Assume that there exists a point $\mathbf{s} \in \{\mathcal{S} \cap \mathcal{S}'\}$. Then, for any $\mathbf{x} \in \mathcal{E}$, the k-nearest neighborhoods function f have the property that*

$$P(f(\mathbf{x}; \mathbf{s}_i|\mathcal{S}) = 1) > P(f(\mathbf{x}; \mathbf{s}_i|\mathcal{S}') = 1),$$

where $f(\mathbf{x}; \mathbf{s}_i|\mathcal{S})$ denotes $f(\mathbf{x}^v; \mathbf{s}_i^v)$ based on the radius of $\mathbf{s}_i^v$ calculated through the sample set $\mathcal{S}^v$. For simplicity, we omit the requisite v in most notations.

Proof. Let $\theta[\mathbf{s}]$ be the hypersphere centered at $\mathbf{s}$, and $\mathbf{y}$ be a point in the ϵ-neighborhood of $\mathbf{s}$, where ϵ is a small positive number such that $\mathbf{y} \in \mathcal{E} \cap \theta[\mathbf{s}]$. Accordingly,

$$\begin{aligned}&P(f(\mathbf{x}; \mathbf{s}|\mathcal{S}) = 1) \\ &\quad = P(\mathbf{x} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}) \times P(\mathbf{s} \in \textit{kNN}(\mathbf{x}) \mid \mathcal{S}) \\ &\quad = P(\mathbf{x} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}) \times P(\mathbf{y} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}) \\ &\quad \times P(\mathbf{s} \in \textit{kNN}(\mathbf{x}) \mid \mathcal{S}) \\ &\quad = P(\mathbf{x}, \mathbf{y} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}) \times P(\mathbf{s} \in \textit{kNN}(\mathbf{x}) \mid \mathcal{S}). \quad (1)\end{aligned}$$

Let $\ell_{\mathbf{xy}}$ denote the Euclidean distance between points $\mathbf{x}$ and $\mathbf{y}$, and let $\mathcal{M}_{\mathcal{D}}(\mathbf{x}; r)$ represent the cardinality of the set of points in $\mathcal{D}$ within a radius r centered at $\mathbf{x}$. Given that $\mathcal{P}_{\mathcal{D}}(w) < \mathcal{P}_{\mathcal{D}'}(w)$ for all $w \in \mathcal{E}$, for any $\mathbf{x} \in \mathcal{E}$, it directly follows that

$$\mathcal{M}_{\mathcal{D}}(\mathbf{x}; \ell_{\mathbf{xs}}) < \mathcal{M}_{\mathcal{D}'}(\mathbf{x}; \ell_{\mathbf{xs}}).$$

Since $\mathcal{S}$ and $\mathcal{S}'$ are random samples from $\mathcal{D}$ and $\mathcal{D}'$, respectively, the expected number of points from $\mathcal{S}$ within a radius $\ell_{\mathbf{xs}}$ is given by

$$\mathbb{E}(\mathcal{M}_{\mathcal{S}}(\mathbf{x}; \ell_{\mathbf{xs}})) = \frac{|\mathcal{S}|}{|\mathcal{D}|}\mathcal{M}_{\mathcal{D}}(\mathbf{x}; \ell_{\mathbf{xs}}).$$

Consequently,

$$\mathbb{E}(\mathcal{M}_{\mathcal{S}}(\mathbf{x}; \ell_{\mathbf{xs}})) < \mathbb{E}(\mathcal{M}_{\mathcal{S}'}(\mathbf{x}; \ell_{\mathbf{xs}})).$$

Now, considering a positive integer k, the probability that $\mathbf{s}$ is among the k-nearest neighbors of $\mathbf{x}$, conditioned on the sampled set $\mathcal{S}$, can be expressed as:

$$P(\mathbf{s} \in \textit{kNN}(\mathbf{x}) \mid \mathcal{S}) \propto \frac{k}{\mathbb{E}(\mathcal{M}_{\mathcal{S}}(\mathbf{x}; \ell_{\mathbf{xs}}))}.$$

Therefore, due to the inverse relationship, we have:

$$P(\mathbf{s} \in \textit{kNN}(\mathbf{x}) \mid \mathcal{S}) > P(\mathbf{s} \in \textit{kNN}(\mathbf{x}) \mid \mathcal{S}'). \quad (2)$$

Moreover, $\mathbf{x}, \mathbf{y} \in \theta[\mathbf{s}]$ if and only if the nearest neighbour of both $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{s}$ in $\mathcal{S}$. Let $\tilde{\mathbf{s}} \in \mathcal{S}$ be the nearest neighbor of $\mathbf{s}$. Then, $\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \max(\ell_{\mathbf{xs}}, \ell_{\mathbf{ys}})$ holds for $\tilde{\mathbf{s}}$ the nearest neighbour of $\mathbf{s}$ in $\mathcal{S}$. Moreover, the triangular inequality $\ell_{\mathbf{xs}} + \ell_{\mathbf{ys}} > \ell_{\mathbf{xy}}$ holds because ℓ is a metric distance. Accordingly,

$$\begin{aligned}&P(\mathbf{x}, \mathbf{y} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S} \subset \mathcal{D}) \\ &= P(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xx}} > \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{ys}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\}) + \\ &\quad P(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xx}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{xy}} - \ell_{\mathbf{ys}} > \ell_{\mathbf{ys}}\}) + \\ &\quad P(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{ys}} > \ell_{\mathbf{xs}}\} \wedge \{\ell_{\mathbf{xs}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{xs}}\}) + \\ &\quad P(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{ys}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{xs}}\} \wedge \{\ell_{\mathbf{xy}} - \ell_{\mathbf{xs}} > \ell_{\mathbf{xs}}\}) \\ &= 2P(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xx}} > \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{ys}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\}) + \\ &\quad 2P(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xx}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{xy}} - \ell_{\mathbf{ys}} > \ell_{\mathbf{ys}}\})\end{aligned} \quad (1)$$

subject to the nearest neighbour $\mathbf{s} \in \mathcal{S}$ of both $\mathbf{x}$ and $\mathbf{y}$. The last equality holds by the symmetry of $\ell_{\mathbf{xs}}$ and $\ell_{\mathbf{ys}}$.

Given a hypersphere $v(c, \ell_{c\mathbf{s}})$ centered at $c \in \mathcal{E}$ and having radius $\ell_{c\mathbf{s}}$ equal to the distance from c to its nearest neighbour $\mathbf{s} \in \mathcal{S}$, let $\mathcal{P}(u(c, \ell_{c\mathbf{s}}))$ be the probability density of probability mass $u(c, \ell_{c\mathbf{s}})$ in $v(c, \ell_{c\mathbf{s}})$; $u(c, \ell_{c\mathbf{s}}) = \int_{v(c,\ell_{c\mathbf{s}})} \mathcal{P}_D(\mathbf{x})\, d\mathbf{x}$. Note that $u(c, \ell_{c\mathbf{s}})$ is strictly monotonic to $\ell_{c\mathbf{s}}$ if $v(c, \ell_{c\mathbf{s}}) \cap \mathcal{X} \neq \varnothing$, since $\mathcal{P}_D$ is strictly positive in $\mathcal{X}$. Then, the followings are derived.

$$\begin{aligned}&P\left(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xs}} > \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{ys}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\}\right) \\ &= P\left(\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xs}} > \ell_{\mathbf{ys}} > \ell_{\mathbf{xy}}/2\right) \\ &= \int_{u(\mathbf{s},\ell_{\mathbf{xy}}/2)}^{1} \mathcal{P}(u(\mathbf{s}, \ell_{\mathbf{s}\tilde{\mathbf{s}}})) \int_{u(\mathbf{x},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{x},\ell_{\mathbf{s}\tilde{\mathbf{s}}})} \mathcal{P}(u(\mathbf{x}, \ell_{\mathbf{xs}})) \times \\ &\quad \int_{u(\mathbf{y},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{y},\ell_{\mathbf{xs}})} \mathcal{P}(u(\mathbf{y}, \ell_{\mathbf{ys}}))\, du(\mathbf{y}, \ell_{\mathbf{ys}})\, du(\mathbf{x}, \ell_{\mathbf{xs}})\, du(\mathbf{s}, \ell_{\mathbf{s}\tilde{\mathbf{s}}}) \\ &\approx \int_{u(\mathbf{s},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})} \mathcal{P}(u(\mathbf{s}, \ell_{\mathbf{s}\tilde{\mathbf{s}}})) \int_{u(\mathbf{x},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{x},\ell_{\mathbf{s}\tilde{\mathbf{s}}})} \mathcal{P}(u(\mathbf{x}, \ell_{\mathbf{xs}})) \times \\ &\quad \int_{u(\mathbf{y},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{y},\ell_{\mathbf{xs}})} \mathcal{P}(u(\mathbf{y}, \ell_{\mathbf{ys}}))\, du(\mathbf{y}, \ell_{\mathbf{ys}})\, du(\mathbf{x}, \ell_{\mathbf{xs}})\, du(\mathbf{s}, \ell_{\mathbf{s}\tilde{\mathbf{s}}}),\end{aligned}$$

$$
\begin{aligned}
&P\left(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{ss}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{xy}} - \ell_{\mathbf{ys}} > \ell_{\mathbf{ys}}\}\right) \\
&= P\left(\{\ell_{\mathbf{s}\tilde{\mathbf{s}}} > \ell_{\mathbf{xs}} > \ell_{\mathbf{xy}} - \ell_{\mathbf{ys}}\} \wedge \{\ell_{\mathbf{xy}}/2) > \ell_{\mathbf{ys}}\}\right) \\
&= \int_{u(\mathbf{s},\ell_{\mathbf{xy}}/2)}^{1} \mathcal{P}(u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})) \int_{0}^{u(\mathbf{y},\ell_{\mathbf{xy}}/2)} \mathcal{P}(u(\mathbf{y},\ell_{\mathbf{ys}}))\times \\
&\quad \int_{u(\mathbf{x},\ell_{\mathbf{xy}}-\ell_{\mathbf{ys}})}^{u(\mathbf{x},\ell_{\mathbf{s}\tilde{\mathbf{s}}})} \mathcal{P}(u(\mathbf{x},\ell_{\mathbf{xs}}))\, du(\mathbf{x},\ell_{\mathbf{xs}})\, du(\mathbf{y},\ell_{\mathbf{ys}})\, du(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}}) \\
&\approx \int_{u(\mathbf{s},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})} \mathcal{P}(u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})) \int_{0}^{u(\mathbf{y},\ell_{\mathbf{xy}}/2)} \mathcal{P}(u(\mathbf{y},\ell_{\mathbf{ys}}))\times \\
&\quad \int_{u(\mathbf{x},\ell_{\mathbf{xy}}/2)}^{u(\mathbf{x},\ell_{\mathbf{s}\tilde{\mathbf{s}}})} \mathcal{P}(u(\mathbf{x},\ell_{\mathbf{xs}}))\, du(\mathbf{x},\ell_{\mathbf{xs}})\, du(\mathbf{y},\ell_{\mathbf{ys}})\, du(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}}),
\end{aligned}
$$

where $\ell_{\mathbf{s}\tilde{\mathbf{s}}} = \sup_{v(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})\subseteq\mathcal{E}} \ell_{\mathbf{s}\tilde{\mathbf{s}}}$. The approximate equality holds by the assumption in the theorem which implies that the integral from $u(\mathbf{s},\hat{\ell}_{\mathbf{s}\mathcal{E}})$ to 1 for $u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})$ is negligible. The same argument applied to $P(\mathbf{x},\mathbf{y} \in \theta[\mathbf{s}'] \mid \mathbf{s}' \in \mathcal{D}' \subset D')$ which derives the identical result.

[Fukunaga, 1990] provided the expressions of $\mathcal{P}(u(c,\ell_{c\mathbf{s}}))$ and $\mathcal{P}(u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}}))$ as

$$
\begin{aligned}
&\mathcal{P}(u(c,\ell_{c\mathbf{s}})) = \psi(1-u(c,\ell_{c\mathbf{s}}))^{\psi-1}, \\
&\mathcal{P}(u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}})) = (\psi-1)(1-u(\mathbf{s},\ell_{\mathbf{s}\tilde{\mathbf{s}}}))^{\psi-2}.
\end{aligned}
$$

With these expressions and the definition of $u(c,\ell_{c\mathbf{s}})$, both $\mathcal{P}(u(c,\ell_{c\mathbf{s}}))$ and $\mathcal{P}(u(\mathbf{s},\ell_{\mathbf{ss}}))$ are lower if $\mathcal{P}_D(\mathbf{s})$ becomes higher. Accordingly,

$$
P(\mathbf{x},\mathbf{y} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}) > P(\mathbf{x},\tilde{\mathbf{s}} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}'). \quad (3)
$$

Thus, based on the equations (1), (2) and (3) the following inequality is derived.

$$
\begin{aligned}
&P(f(\mathbf{x};\mathbf{s}|\mathcal{S}) = 1) \\
&\quad > P(\mathbf{x},\tilde{\mathbf{s}} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}') \times P(s \in \mathit{kNN}(\mathbf{x}) \mid \mathcal{S}') \\
&\quad = P(\mathbf{x} \in \theta[\mathbf{s}] \mid \mathbf{s} \in \mathcal{S}') \times P(s \in \mathit{kNN}(\mathbf{x}) \mid \mathcal{S}') \\
&\quad = P(f(\mathbf{x};\mathbf{s}_i|\mathcal{S}') = 1).
\end{aligned}
$$

This result proves the theorem.

□

References

[Fukunaga, 1990] Keinosuke Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.