## 8 Appendix - Lemmas and Derivations

### 8.1 Lemmas

Lemma 8.1. Suppose a probability distribution $P\left(\mathbf{x}_{i}, \mathbf{x}_{y}\right)=f\left(\mathbf{x}_{i}, \mathbf{x}_{j}\right)$ is defined in terms of $\mathbf{x}_{i}=\left(x_{i 1}, x_{i 2}, \ldots, x_{i k}\right) \in \mathbb{R}^{k}$ and $\mathbf{x}_{j}=\left(x_{j 1}, x_{j 2}, \ldots, x_{j k}\right) \in \mathbb{R}^{k}$. Let $\mathbf{m}=\frac{\mathbf{x}_{i}+\mathbf{x}_{j}}{2}$ be the midpoint of $\mathbf{x}_{i}$ and $\mathbf{x}_{j}$, and let $\mathbf{d}=\frac{\mathbf{x}_{i}-\mathbf{x}_{j}}{2}$ be the points' (symmetric) displacement from the midpoint. Then $P(\mathbf{m}, \mathbf{d})=2^{k} f\left(\mathbf{x}_{i}, \mathbf{x}_{j}\right)$.

Likewise, the factor is $\frac{1}{2^{k}}$ for the inverse transformation: $P\left(\mathbf{x}_{i}, \mathbf{x}_{j}\right)=\frac{1}{2^{k}} P(\mathbf{m}, \mathbf{d})$.
Proof. The transformation is one-to-one. For each vector component $l$, the transforms and their inverses are:

$$
\begin{aligned}
m_{l} & =\frac{x_{i l}+x_{j l}}{2} & x_{i l} & =m_{l}+d_{l} \\
d_{l} & =\frac{x_{i l}-x_{j l}}{2} & x_{j l} & =m_{l}-d_{l}
\end{aligned}
$$

We can show by induction that in $k$ dimensions, the determinant of the Jacobian, $\left|A_{k}\right|$, is $-2^{k}$. Then, its absolute value $2^{k}$ is the factor used in the transformation.

For a single dimension, $\left|A_{1}\right|$ is

$$
\left|\begin{array}{ll}
\frac{\partial x_{i l}}{\partial m_{l}}=1 & \frac{\partial x_{i l}}{\partial d_{l}}=1 \\
\frac{\partial x_{j l}}{\partial m_{l}}=1 & \frac{\partial x_{j l}}{\partial d_{l}}=-1
\end{array}\right|=-2=-2^{1} .
$$

In general,

$$
A_{k}=\left[\begin{array}{ccccccc}
\frac{\partial x_{i 1}}{m_{1}}=1 & \frac{\partial x_{i 1}}{\partial d_{1}}=1 & \frac{\partial x_{i 1}}{\partial m_{2}}=0 & \frac{\partial x_{i 1}}{\partial d_{1}}=0 & \cdots & \frac{\partial x_{i 1}}{\partial m_{1}}=0 & \frac{\partial x_{i 1}}{\partial d_{k}}=0 \\
\frac{\partial x_{j 1}}{\partial m_{1}}=1 & \frac{\partial x_{j 1}}{\partial d_{1}}=-1 & \frac{\partial x_{j 1}}{\partial m_{2}}=0 & \frac{\partial x_{j 1}}{\partial d_{21}}=0 & \cdots & \frac{\partial x_{j 1}}{\partial m_{k}}=0 & \frac{\partial x_{j 1}}{\partial d_{k}}=0 \\
\frac{\partial x_{i 2}}{\partial m_{1}}=0 & \frac{\partial x_{i 2}}{\partial d_{1}}=0 & \frac{\partial x_{2}}{\partial m_{2}}=1 & \frac{\partial x_{i 2}}{\partial d_{2}}=1 & \cdots & \frac{\partial x_{2}}{\partial m_{k}}=0 & \frac{\partial x_{i 2}}{\partial d_{k}}=0 \\
\frac{\partial x_{j 2}}{\partial m_{1}}=0 & \frac{\partial x_{2}}{\partial d_{1}}=0 & \frac{\partial x_{j 2}}{\partial m_{2}}=1 & \frac{\partial x_{j 2}}{\partial d_{2}}=-1 & \cdots & \frac{\partial x_{j 2}}{\partial m_{k}}=0 & \frac{\partial x_{j 2}}{\partial d_{k}}=0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{\partial x_{i k}}{\partial m_{1}}=0 & \frac{\partial x_{i k}}{\partial d_{1}}=0 & \frac{\partial x_{i k}}{\partial m_{2}}=0 & \frac{\partial x_{i k}}{\partial d_{2 k}}=0 & \cdots & \frac{\partial x_{i k}}{\partial m_{k}}=1 & \frac{\partial x_{i k}}{\partial d_{k}}=1 \\
\frac{\partial x_{j k}}{\partial m_{1}}=0 & \frac{\partial x_{k}}{\partial d_{1}}=0 & \frac{\partial x_{j k}}{\partial m_{2}}=0 & \frac{\partial x_{j k}}{\partial d_{2}}=0 & \cdots & \frac{x_{j k}}{\partial m_{k}}=1 & \frac{\partial x_{j 1}}{\partial d_{k}}=-1
\end{array}\right]
$$

$$
\left|A_{k}\right|=\left|\begin{array}{ccccccc}
1 & 1 & 0 & 0 & \cdots & 0 & 0 \\
1 & -1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & & & \\
0 & 0 & & & & \\
\vdots & \vdots & & A_{k}-1 \\
0 & 0 & & \\
0 & 0 & &
\end{array}\right|=(-1)\left|A_{k-1}\right|-(1)\left|A_{k-1}\right|=-2\left(-2^{k-1}\right)=-2^{k}
$$

Lemma 8.2. Given points $\mathbf{x}_{i}$ and $\mathbf{x}_{j} \in \mathbb{R}^{k}$, with midpoint $\mathbf{m}_{i j}$ and displacement vector $\mathbf{d}_{i j}$ from the midpoint. Define $m=\left\|\mathbf{m}_{i j}-\boldsymbol{\mu}\right\|$ and $d=\left\|\mathbf{d}_{i j}\right\|$. Then $\left\|\mathbf{x}_{i}-\boldsymbol{\mu}\right\|^{2}+\left\|\mathbf{x}_{j}-\boldsymbol{\mu}\right\|^{2}=$ $2\left(m^{2}+d^{2}\right)$.
It then follows (see main text, Eqs. (7) and (8)) that if $\phi \sim \operatorname{Normal}\left(\boldsymbol{\mu}, \sigma^{2} I\right)$, then $P\left(\mathbf{x}_{i} \mid\right.$ $\phi) P\left(\mathbf{x}_{j} \mid \phi\right)$ depends on $\mathbf{x}_{i}$ and $\mathbf{x}_{j}$ only through $m$ and d.

Proof. The vectors $\left(\boldsymbol{\mu}, \mathbf{x}_{i}\right)$ and ( $\left.\boldsymbol{\mu}, \mathbf{x}_{j}\right)$ define a plane, which contains $\mathbf{m}_{i j}$. In that plane, Figure 4 is as shown. The law of cosines tells us that $\left\|\mathbf{x}_{i}-\boldsymbol{\mu}\right\|^{2}=m^{2}+d^{2}-2 m d \cos f$, and that $\left\|\mathbf{x}_{j}-\boldsymbol{\mu}\right\|^{2}=m^{2}+d^{2}-2 m d \cos (\pi-f)$. Since $\cos (\pi-f)=-\cos f$, we can rewrite $\left\|\mathbf{x}_{j}-\boldsymbol{\mu}\right\|^{2}=m^{2}+d^{2}+2 m d \cos f$. This yields $\left\|\mathbf{x}_{i}-\boldsymbol{\mu}\right\|^{2}+\left\|\mathbf{x}_{j}-\boldsymbol{\mu}\right\|^{2}=2\left(m^{2}+d^{2}\right)$.


Figure 4: Triangle formed by $\mathbf{x}_{i}, \mathbf{x}_{j}$, and $\boldsymbol{\mu}$.

Lemma 8.3. Given $\mathbf{z} \in \mathbb{R}^{k}$ with magnitude $z=\|\mathbf{z}\|$, and with a probability density that depends only on that magnitude: $f_{\mathbf{z}}(\mathbf{z})=g(z)$. Then, changing variables to write the density as a function of $z$ gives $f_{z}(z)=g(z) \frac{2 z^{k-1} \pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}$.

Explanation. To transform, we would write $\mathbf{z}$ in spherical coordinates $\left(\mathbf{z}=\left(z, \alpha_{1}, \alpha_{2}, \ldots, \alpha_{k-1}\right)\right.$, with $\alpha_{1}, \ldots, \alpha_{k-2} \in[0, \pi]$ and $\left.\alpha_{k-1} \in[0,2 \pi)\right)$, then integrate out the angles. The factor introduced by this process is $S_{k}=\frac{2 z^{k-1} \pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}$, the surface area of a hypersphere of radius $z$ in $\mathbb{R}^{k}$. (References for the formula: http://mathworld.wolfram.com/Hypersphere.html and http://en.wikipedia.org/wiki/Hypersphere.)

Lemma 8.4. If $f_{y}(y)=\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \frac{y^{k-1}}{a^{k}} e^{\frac{-y^{2}}{2 a^{2}}}$, defined on $y \geq 0$ with $a \in \mathbb{R}$ and $k \in \mathbb{Z}, k>0$, then $z=\frac{y}{a} \sim \chi_{k}$.

Proof. Perform the change of variables: $z=\frac{y}{a}$, so $y=a z$ and $\frac{d y}{d z}=a$.

$$
\begin{align*}
f_{y}(y) & =\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \frac{y^{k-1}}{a^{k}} e^{\frac{-y^{2}}{2 a^{2}}}  \tag{14}\\
f_{z}(z) & =\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} z^{k-1}\left(\frac{1}{a}\right) e^{\frac{-z^{2}}{2}}\left|\frac{d y}{d z}\right|  \tag{15}\\
& =\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} z^{k-1} e^{\frac{-z^{2}}{2}} \tag{16}
\end{align*}
$$

Equation (16) is exactly the density function for the distribution $\chi_{k}$. We will also use the notation $\chi_{k}(z)$ to represent that density function, so we would write above that $f_{z}(z)=$ $\chi_{k}(z)$ and also that $f_{y}(y)=\chi_{k}\left(\frac{y}{a}\right)\left(\frac{1}{a}\right)$.

### 8.2 Distributions for positive and negative pairs

Beginning with the expression from (6) for the density function for positive pairs,

$$
\begin{equation*}
P\left(\mathbf{m}_{i j} \mid \phi\right) P\left(\mathbf{d}_{i j} \mid \epsilon\right)=\left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{k} e^{-\frac{m^{2}}{2 \sigma^{2}}}\left(\frac{1}{\sqrt{2 \pi} \nu}\right)^{k} e^{-\frac{d_{j i}^{2}}{2 \nu^{2}}}, \tag{17}
\end{equation*}
$$

we apply Lemma 8.3 to the first and second parts, respectively, to get

$$
\begin{align*}
P(m \mid \phi) P(d \mid \epsilon) & =\left(\frac{2 m^{k-1} \pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right)\left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{k} e^{-\frac{m^{2}}{2 \sigma^{2}}}\left(\frac{2 d^{k-1} \pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right)\left(\frac{1}{\sqrt{2 \pi} \nu}\right)^{k} e^{-\frac{d_{i j}^{2}}{2 \nu^{2}}}  \tag{18}\\
& =\left(\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right) \frac{m^{k-1}}{\sigma^{k}} e^{-\frac{m^{2}}{2 \sigma^{2}}}\left(\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right) \frac{d^{k-1}}{\nu^{k}} e^{-\frac{d^{2}}{2 \nu^{2}}} . \tag{19}
\end{align*}
$$

With the use of Lemma 8.4, we can recognize this as the product of two $\chi_{k}$ distributions: $\frac{m}{\sigma}=m^{\prime} \sim \chi_{k}$, and $\frac{d}{\nu}=\frac{d^{\prime}}{t} \sim \chi_{k}$. Rewriting in terms of $m^{\prime}, d^{\prime}$ and $t$, and using the notation $\chi_{k}(z)$ to represent the density function for $\chi_{k}$ :

$$
\begin{equation*}
P\left(m^{\prime} \mid \phi\right) P\left(d^{\prime} \mid \epsilon\right)=\chi_{k}\left(m^{\prime}\right) \chi_{k}\left(\frac{d^{\prime}}{t}\right)\left(\frac{1}{t}\right) \tag{20}
\end{equation*}
$$

This is intuitively reasonable because $\chi_{k}$ is known to describe the distance from the origin to points that are distributed as $k$-dimensional normals, which is exactly where $\mathbf{m}_{i j}$ and $\mathbf{d}_{i j}$ came from.

For the negative density, we take the expression from Equation (8) and transform from coordinates $\left(\mathbf{x}_{i}, \mathbf{x}_{j}\right)$ to $\left(\mathbf{m}_{i j}, \mathbf{d}_{i j}\right)$ to $(m, d)$.

$$
\begin{align*}
P\left(\mathbf{x}_{i} \mid \phi\right) P\left(\mathbf{x}_{j} \mid \phi\right) & =\left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{2 k} e^{-\frac{m^{2}+d^{2}}{\sigma^{2}}}  \tag{21}\\
P\left(\mathbf{m}_{i j} \mid \phi\right) P\left(\mathbf{d}_{j j} \mid \phi\right) & =2^{k}\left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{k} e^{\frac{-m^{2}}{\sigma^{2}}}\left(\frac{1}{\sqrt{2 \pi} \sigma}\right)^{k} e^{\frac{-d^{2}}{\sigma^{2}}}  \tag{22}\\
P(m \mid \phi) P(d \mid \phi) & =\left(\frac{2 m^{k-1} \pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right)\left(\frac{1}{\sqrt{\pi} \sigma}\right)^{k} e^{\frac{-m^{2}}{\sigma^{2}}}\left(\frac{2 d^{k-1} \pi^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right)\left(\frac{1}{\sqrt{\pi} \sigma}\right)^{k} e^{\frac{-d^{2}}{\sigma^{2}}}  \tag{23}\\
& =\left(\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right)\left(\frac{\sqrt{2}}{\sigma}\right)^{k} m^{k-1} e^{\frac{-m^{2}}{\sigma^{2}}}\left(\frac{2^{1-\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)}\right)\left(\frac{\sqrt{2}}{\sigma}\right)^{k} d^{k-1} e^{\frac{-d^{2}}{\sigma^{2}}} \tag{24}
\end{align*}
$$

Lemma 8.4 now applies, to show this term is also a product of two $\chi_{k}$ distributions: $\frac{m \sqrt{2}}{\sigma} \sim \chi_{k}$, and $\frac{d \sqrt{2}}{\sigma} \sim \chi_{k}$, or in terms of $m^{\prime}$ and $d^{\prime}:$

$$
\begin{equation*}
P\left(m^{\prime} \mid \phi\right) P\left(d^{\prime} \mid \phi\right)=2 \chi_{k}\left(m^{\prime} \sqrt{2}\right) \chi_{k}\left(d^{\prime} \sqrt{2}\right) \tag{26}
\end{equation*}
$$

### 8.3 The effect of increasing $k$, the number of dimensions

Figure 2 was for 2-dimensional data. In $k$ dimensions the contour lines have the same shape for a given value of $t$, but as $k$ grows the distributions end up better separated.
The peak of the negatives is always at $\left(\frac{\sqrt{k-1}}{\sqrt{2}}, \frac{\sqrt{k-1}}{\sqrt{2}}\right)$, and that of the positives is at $(\sqrt{k-1}, t \sqrt{k-1})$. As the dimensionality varies, the relative positions of the peaks are the same apart for the scale factor $\sqrt{k-1}$. However, the variances of the distributions do not scale as fast. So, compared to their positions, the distributions get proportionally narrower and easier to distinguish.


Figure 5: Theoretical distributions of positive and negative pairs, as functions of parameters, and score assigned. From left to right, $t$ takes on values (.02, .3, .7, 1). From top to bottom, number of dimensions $k=1,2,10,100 . n=25, \mathrm{E}(r)=10$.

