null_discriminability

Eric Bridgeford

January 17, 2019

Setup

Given the following two-class problem:

$$X_i \overset{d}{\sim} \mathcal{N}\left(\mu_i^{(x)}, \Sigma_i^{(x)}\right)$$

with $\mu_i^{(x)} \in \mathbb{R}^p, \Sigma_i^{(x)} \in \mathbb{R}^{p \times p}$, then the goal is the distribution of $\|Y\| = \|X_1 - X_2\|$, where:

$$Y = X_1 - X_2 \overset{d}{\sim} \mathcal{N}\left(\mu^{(y)} = \mu_1^{(x)} - \mu_2^{(x)}, \Sigma^{(y)} = \Sigma_1^{(x)} + \Sigma_2^{(x)}\right)$$

To simplify the math, assume that $\Sigma_{i,j}^{(x)} = \sigma_i^{(x)}$ when $i = j$, and 0 otherwise. That is, observations between dimensions are assumed to be independent. Let $Y = (Y_i)_{i=1}^p$, so Y_i is a random variable for each dimension, where $\Sigma_{i,j}^{(y)} = \sigma_i^{(y)}$ when $i = j$ and zero otherwise. Then:

$$\|Y\|^2 = \sum_{i=1}^p Y_i$$

and:

$$\|Y\| = \sqrt{\sum_{i=1}^p Y_i}$$

Theoretical Approach

Approach 1

Note that for $A = Z_1 + \mu$ where $A \overset{d}{\sim} \mathcal{N}(\mu, 1)$ and Z_1 is the standard gaussian, we can find the MGF:

$$\begin{aligned}
M_A(t) = \mathbb{E}\left[e^{tA^2}\right] &= \mathbb{E}\left[e^{t(Z+\mu)^2}\right] \\
&= \int_{-\infty}^{\infty} e^{t(z+\mu)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, \mathrm{d}z \\
&= \int_{-\infty}^{\infty} e^{t(z+\mu)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, \mathrm{d}z
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\left(\frac{1}{2}-t\right)z^2 + 2\mu t z + \mu^2 t\right) \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\left(\frac{1}{2}-t\right)\left(z - \frac{2\mu t}{1-2t}\right)^2 + \mu^2 t + \frac{2\mu^2 t^2}{1-2t}\right) \mathrm{d}z \\
&= \frac{\sqrt{1-2t}}{\sqrt{2\pi}\sqrt{1-2t}} \exp\left(\mu^2 t + \frac{2\mu^2 t^2}{1-2t}\right) \int_{-\infty}^{\infty} \exp\left(-\left(\frac{1}{2}-t\right)\left(z - \frac{2\mu t}{1-2t}\right)^2\right) \mathrm{d}z \\
&= \frac{\sqrt{1-2t}}{\sqrt{2\pi}\sqrt{1-2t}} \exp\left(\mu^2 t + \frac{2\mu^2 t^2}{1-2t}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{\left(z - \frac{2\mu t}{1-2t}\right)^2}{2(1-2t)^{-1}}\right) \mathrm{d}z \\
&= \frac{1}{\sqrt{1-2t}} \exp\left(\mu^2 t + \frac{2\mu^2 t^2}{1-2t}\right) \frac{\sqrt{1-2t}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(z - \frac{2\mu t}{1-2t}\right)^2 (1-2t)\right) \mathrm{d}z \\
&= \frac{1}{\sqrt{1-2t}} \exp\left(\mu^2 t + \frac{2\mu^2 t^2}{1-2t}\right) \qquad \text{pdf of Gaussian} \\
&= \frac{1}{\sqrt{1-2t}} \exp\left(\frac{\mu^2 t}{1-2t}\right)
\end{aligned}
$$

which is the MGF of the non-central chi-squared distribution with non-centrality parameter μ^2. Then:

$$
\left(\frac{Y_i}{\sigma_{y,i}}\right)^2 \overset{d}{\sim} \chi_1^2\left(\lambda_i^{(y)}\right),\ \lambda_i^{(y)} = \left(\frac{\mu_i^{(y)}}{\sigma_i^{(y)}}\right)^2
$$

is non-central chi-squared with 1 degree of freedom and non-centrality parameter $\lambda_i^{(y)} = \left(\frac{\mu_i^{(y)}}{\sigma_i^{(y)}}\right)^2$. Then we can write:

$$
\|Y\|^2 = \sum_{i=1}^{p} \sigma_i^2 \left(\frac{Y_i}{\sigma_i^{(y)}}\right)^2
$$

which is a linear combination of non-central chi-squareds, which is the definition of the generalized chi-squared distribution.

Approach 2

Let $Z_p \overset{d}{\sim} \mathcal{N}(0, I_p)$, the standard isotropic gaussian with p dimensions. Then by definition with $Y \overset{d}{\sim} \mathcal{N}\left(\mu^{(y)}, \Sigma^{(y)}\right)$, since $\Sigma^{(y)} \succ 0$, Σ has the eigendecomposition $\Sigma = Q'\Lambda Q$ where Q is the orthonormal matrix of eigenvectors and Λ the square diagonal matrix of positive eigenvalues, where $\Lambda = W'W$ and $W = \Lambda^{\frac{1}{2}}$. Then $\Sigma = (WQ)'(WQ) = L'L$ is a cholesky decomposition of Σ. Then it is clear that $Y = LZ_p + \mu^{(y)}$, so:

$$
\begin{aligned}
\|Y\|^2 = Y'Y &= \left(LZ_p + \mu^{(y)}\right)'\left(LZ_p + \mu^{(y)}\right) \\
&= \left(Z_p + L^{-1}\mu^{(y)}\right)'\Sigma\left(Z_p + L^{-1}\mu^{(y)}\right) \\
&= \left(Z_p + L^{-1}\mu^{(y)}\right)'\Sigma\left(Z_p + L^{-1}\mu^{(y)}\right)
\end{aligned}
$$

Then with $S = QZ_p$, S is also multivariate normal since it is a full-rank linear transformation of Z_p (recall Q is full-rank by definition of the eigendecomposition of $\Sigma \succ 0$). Note also that it is $S \overset{d}{\sim} \mathcal{N}(0, I_p)$ since $Q'Q = QQ' = I_p$ by definition of orthonormal. Then it is clear that:

$$\begin{aligned}
Y'Y &= \left(Z_p + L^{-1}\mu^{(y)}\right)' Q'\Lambda Q\left(Z_p + L^{-1}\mu^{(y)}\right) \\
&= \left(QZ_p + QL^{-1}\mu^{(y)}\right)' \Lambda\left(QZ_p + QL^{-1}\mu^{(y)}\right) \\
&= \left(S + QL^{-1}\mu^{(y)}\right)' \Lambda\left(S + QL^{-1}\mu^{(y)}\right) \\
&= (S+u)'\Lambda(S+u) \\
&= \sum_{i=1}^{p} (S_i + u_i)^2 \lambda_i
\end{aligned}$$

where $u = QL^{-1}\mu^{(y)}$. Then $u_i = Q\frac{\mu^{(y)}}{\sigma_i^{(y)}}$, since $L_{i,i} = \sigma_i^{(y)}$ as Σ is diagonal and 0 otherwise. Note then that by definition, $(S_i + u_i)^2$ is non-central chi-squared distributed, so again we see that $\|Y\|^2$ is a linear combination of independent non-central chi-squareds.

Then by both of the above approaches, it is clear that $\|Y\|$ is distributed as the generalized chi-distribution. In the case when the variances are the same, this is simply the Rice distribution.

I'm honestly shocked but apparently when Σ is not a simple scaling of the identity, this distribution does not have any real empirical methods associated with it (ie, nice pdf, cdf, etc).

Empirical Approach

Algorithm 1: Empirical Discriminability

Data: $\mu_i \in \mathbb{R}^p$ the means of our 2 classes.
$\Sigma_i \in \mathbb{R}^{p \times p}$ the covariances of our classes.
n_i the number of elements in each class.
n the number of realizations to empirically sample, per-class.
Result: D the theoretical discriminability using an empirical approximation.

```
for l = 1 : n do
    for i = 1 : p do
        for j = 1 : p do
            sample x_i from N(μ_i, Σ_i).
            sample x_j from N(μ_j, Σ_j).
            Let d_{l,i,j} = ||x_i − x_j||
        end
    end
end
for i = 1 : p do
    for j = 1 : p do
        Let P_{i,j} be the relative mass of d_{:,i,i} < d_{:,i,j}.
    end
end
Let D = Σ_{i=1}^p π_i π_j P_{i,j}, where π_i = n_i / n.
```