Purity calculation

Oct 05, 2019

1 Background

1.1 Kullback-Leibler divergence

In statistics, the Kullback-Leibler (KL) divergence is a measure of how one probability distribution F_1 is different from a sceond reference probability distribution F_2. We may apply the KL divergence idea in the clustering problem, since the larger the KL divergence between distributions, the more pure the groups/clusters are.

For distributions F_1 and F_2 of a continuous random variable, the KL divergence is defined as:

$$D_{KL}(F_1||F_2) = \int_{-\infty}^{+\infty} f_1(x)log(\frac{f_1(x)}{f_2(x)})dx \tag{1}$$

where f_1 and f_2 denote the probability density of F_1 and F_2.

Besides, $log(x) \leq x - 1$ is always true, then

$$\begin{aligned}\int -log(\frac{f_2(x)}{f_1(x)})f_1(x)dx &\geq \int -(\frac{f_2(x)}{f_1(x)} - 1)f_1(x)dx \\ &= \int [f_2(x) - f_1(x)]dx = 0\end{aligned}$$

The $D_{KL}(F_1||F_2)$ is always bigger or equal to than 0. Similarly, the $D_{KL}(F_2||F_1)$ is also always bigger or equal to than 0.

1.2 Application

In our setting, we assume the outcomes are from a linear mixed model:

$$\boldsymbol{Y} = \boldsymbol{S}(\boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\Gamma}(\boldsymbol{\alpha}'\boldsymbol{x})) + \boldsymbol{\epsilon}. \tag{2}$$

where,

- $\mathbf{S}$ is the matrix of times (intercept, linear, and quadratic term)
- $\boldsymbol{\beta}$ is the vector of covariates for fixed effects of $\boldsymbol{S}$
- $\boldsymbol{b}$ is the vector of random effects

- $\boldsymbol{\Gamma}$ is the vector of fixed effects of the baseline covariates.
- $\boldsymbol{\alpha}'\boldsymbol{x}$ is the combination of the input baseline covariates.
- $\boldsymbol{\alpha}$ has the restriction that $||\boldsymbol{\alpha}|| = 1$

Define the covariate matrix of $\boldsymbol{S}$ as $\boldsymbol{z}$. The $\boldsymbol{z}$ contains both fixed effects and random effects.

$$\boldsymbol{z} = \boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\Gamma}\boldsymbol{w}$$

That is, we have distributions for the mixed-effect model coefficients $\boldsymbol{z}$ given $w = \boldsymbol{\alpha}'\boldsymbol{x}$, where

$$\boldsymbol{z}|w \sim N(\boldsymbol{\beta}_j + \boldsymbol{\Gamma}_j w, \boldsymbol{D}_j),$$

for treatment $j = 1, 2$. Besides, we assume the baseline biosignature x follows distribution with mean μ_x and covariance matrix Σ_x

Based on the Kullback-Leibler divergence, we define the *purity* of the data, which represents how much the differences between the treatment group distribution $f_1(x)$ and the placebo group distribution $f_2(x)$. We define the **purity function** regard to a subject with baseilne biosignature $\boldsymbol{x}$ (i.g. the **purity function**given α and the baseline biosignature $\boldsymbol{x}$) as:

$$\begin{aligned} g(\boldsymbol{\alpha}'\boldsymbol{x}) =& D_{KL}(F_1||F_2) + D_{KL}(F_2||F_1) \\ =& \int log(f_1(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x}))f_1(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x})dz - \int log(f_2(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x}))f_1(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x})dz \\ &+ \int log(f_2(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x}))f_2(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x})dz - \int log(f_1(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x}))f_2(\boldsymbol{z}|\boldsymbol{\alpha}'\boldsymbol{x})dz \end{aligned} \tag{3}$$

where,

$$\begin{aligned} f_1(\boldsymbol{z}|\boldsymbol{w}) &= \frac{1}{\sqrt{((2\pi)^p|\boldsymbol{D}_1|)}} exp(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_1)'\boldsymbol{D}_{\boldsymbol{1}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_1)) \\ f_2(\boldsymbol{z}|\boldsymbol{w}) &= \frac{1}{\sqrt{((2\pi)^p|\boldsymbol{D}_2|)}} exp(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_2)) \\ \boldsymbol{\mu}_1 &= \boldsymbol{\beta}_1 + \boldsymbol{\Gamma}_1 w, \boldsymbol{\mu}_2 = \boldsymbol{\beta}_2 + \boldsymbol{\Gamma}_2 w \end{aligned}$$

Furthermore, we define f_w as the distribution of the combination of baseline signature, $w = \boldsymbol{\alpha}'\boldsymbol{x}$.

Then the purity function regards to the whole data set is defined as:

$$\begin{aligned} \text{purity}(\boldsymbol{\alpha}) =& \int g(\boldsymbol{\alpha}'\boldsymbol{x})f_w(\boldsymbol{\alpha}'\boldsymbol{x})d\boldsymbol{\alpha}'\boldsymbol{x} \\ =& E(g(\boldsymbol{\alpha}'\boldsymbol{x})) \end{aligned} \tag{4}$$

Therefore, we may estimate the dataset's purity given a vector α by the mean value of $g()$ function,

$$\hat{\text{purity}}(\boldsymbol{\alpha}) = \bar{g}(\boldsymbol{\alpha}'\boldsymbol{x})$$

1.2.1 Purity Calculation

We can separate Equation(3) into four parts: $\int f_1 log f_1$, $\int f_2 log f_2$, $\int f_1 log f_2$, and $\int f_2 log f_1$.

- For $\int f_1 log f_1$ and $\int f_2 log f_2$:

$$\begin{aligned}\int f_1 log f_1 =& E_1(log(f_1))\\ =& E_1(-\frac{p}{2}log(2\pi) - \frac{1}{2}log(|\boldsymbol{D}_1|) - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1))\\ =& -\frac{p}{2}log(2\pi) - \frac{1}{2}log(|\boldsymbol{D}_1|) - \frac{1}{2}E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)]\end{aligned}$$

And

$$\begin{aligned}E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)] =& E_1[tr((\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1))]\\ =& E_1[tr(\boldsymbol{D}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)'(\boldsymbol{z}-\boldsymbol{\mu}_1))]\\ =& tr(E_1[\boldsymbol{D}_1^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)'(\boldsymbol{z}-\boldsymbol{\mu}_1)])\\ =& tr(\boldsymbol{D}_1^{-1}E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1)'(\boldsymbol{z}-\boldsymbol{\mu}_1)])\\ =& tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_1) = tr(\boldsymbol{I}_p) = p\end{aligned}$$

Therefore,

$$\int f_1 log f_1 = -\frac{p}{2}log(2\pi) - \frac{1}{2}log(|\boldsymbol{D}_1|) - \frac{p}{2} \tag{5}$$

Similarly,

$$\int f_2 log f_2 = -\frac{p}{2}log(2\pi) - \frac{1}{2}log(|\boldsymbol{D}_2|) - \frac{p}{2} \tag{6}$$

- For $\int f_1 log f_2$ and $\int f_2 log f_1$

$$\begin{aligned}\int f_1 log f_2 =& E_1(log f_2)\\ =& E_1(-\frac{p}{2}log(2\pi) - \frac{1}{2}log(|\boldsymbol{D}_2|) - \frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_2))\\ =& -\frac{p}{2}log(2\pi) - \frac{1}{2}log(|\boldsymbol{D}_2|) - \frac{1}{2}E_1[(\boldsymbol{z}-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_2)]\end{aligned}$$

And

$$
\begin{aligned}
E_1[(\boldsymbol{z}-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(z-\boldsymbol{\mu}_2)] =& E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1+\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1+\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)]\\
=& E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)\\
&+(\boldsymbol{z}-\boldsymbol{\mu}_1)\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)]\\
=& E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)]+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}E_1(\boldsymbol{z}-\boldsymbol{\mu}_1)+\\
&E_1(\boldsymbol{z}-\boldsymbol{\mu}_1)')D_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
=& E_1[(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)]+0+0+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
=& E_1[tr(\boldsymbol{z}-\boldsymbol{\mu}_1)'\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1))]+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
=& E_1[tr(\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)'(\boldsymbol{z}-\boldsymbol{\mu}_1))]+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
=& tr(E_1[\boldsymbol{D}_2^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_1)'(\boldsymbol{z}-\boldsymbol{\mu}_1)])+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
=& tr(\boldsymbol{D}_2^{-1}E_1[(z-\boldsymbol{\mu}_1)'(\boldsymbol{z}-\boldsymbol{\mu}_1)])+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
=& tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)
\end{aligned}
$$

Therefore,

$$
\int f_1 log f_2 = -\frac{p}{2}log(2\pi)-\frac{1}{2}log(|\boldsymbol{D}_2|)-\frac{1}{2}\big(tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big) \tag{7}
$$

Similarly,

$$
\int f_2 log f_1 = -\frac{p}{2}log(2\pi)-\frac{1}{2}log(|\boldsymbol{D}_1|)-\frac{1}{2}\big(tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_1^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big) \tag{8}
$$

Therefore, the equation (1) is:

$$
(3)=(5)-(7)+(6)-(8)
$$

That is,

$$
\begin{aligned}
&\int log(f_1)f_1-\int log(f_2)f_1+\int log(f_2)f_2-\int log(f_1)f_2\\
&=\big(-\frac{p}{2}log(2\pi)-\frac{1}{2}log(|\boldsymbol{D}_1|)-\frac{p}{2}\big)\\
&-\big(-\frac{p}{2}log(2\pi)-\frac{1}{2}log(|\boldsymbol{D}_2|)-\frac{1}{2}\big(tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big)\big)\\
&+\big(-\frac{p}{2}log(2\pi)-\frac{1}{2}log(|\boldsymbol{D}_2|)-\frac{p}{2}\big)\\
&-\big(-\frac{p}{2}log(2\pi)-\frac{1}{2}log(|\boldsymbol{D}_1|)-\frac{1}{2}\big(tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)+(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{D}_1^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\big)\big)\\
&=-p+\frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+\frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)+\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)
\end{aligned}
$$

where $\boldsymbol{\mu}_1=\boldsymbol{\beta}_1+\boldsymbol{\Gamma}_1\boldsymbol{\alpha}'\boldsymbol{x}$, $\boldsymbol{\mu}_2=\boldsymbol{\beta}_2+\boldsymbol{\Gamma}_2\boldsymbol{\alpha}'\boldsymbol{x}$.

Besides,

$$\begin{aligned}
&(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\\
&=\big(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2+(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\boldsymbol{\alpha}'\boldsymbol{x}\big)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})\big(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2+(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\boldsymbol{\alpha}'\boldsymbol{x}\big)\\
&=(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\\
&+2\big[(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\boldsymbol{x}'\boldsymbol{\alpha}\\
&+\boldsymbol{\alpha}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\alpha}\big((\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\big)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\big)
\end{aligned}$$

Therefore, the purity for a subject with baseline biosignature x is:

$$\begin{aligned}
g(\boldsymbol{\alpha}'\boldsymbol{x})=&-p+\frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+\frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)\\
&+\frac{1}{2}\big\{(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\\
&+2\big[(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\boldsymbol{x}'\boldsymbol{\alpha}\\
&+\boldsymbol{\alpha}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\alpha}\big((\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\big)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\big)\big\}
\end{aligned}\tag{9}$$

The dataset′s purity, which is the expectation of the $g()$ function is:

$$\begin{aligned}
\text{purity}(\alpha)=&E(g(\alpha'x))\\
=&-p+\frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+\frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)\\
&+\frac{1}{2}\big\{A_1+2A_2E(\boldsymbol{x}'\boldsymbol{\alpha})+A_3E(\boldsymbol{\alpha}'\boldsymbol{x}\boldsymbol{x}'\boldsymbol{\alpha})\big\}\\
=&A_0+\frac{A_1}{2}+A_2\boldsymbol{\mu}_x'\boldsymbol{\alpha}+\frac{A_3}{2}[tr(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha})+\boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}]\\
=&A_0+\frac{A_1}{2}+A_2\boldsymbol{\mu}_x'\boldsymbol{\alpha}+\frac{A_3}{2}[\alpha'\boldsymbol{\Sigma}_x\boldsymbol{\alpha}+\boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}]
\end{aligned}\tag{10}$$

where

- $A_0=-p+\frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+\frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)$
- $A_1=(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)$
- $A_2=(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)$
- $A_3=(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\big)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)$

All A_0, A_1, A_2, A_3 are scalars.

Therefore, if the distribution of x, f_1, f_2 and $\boldsymbol{\alpha}$ are known, we can calculate the purity by:

$$\text{purity}(\boldsymbol{\alpha})=A_0+\frac{A_1}{2}+A_2\boldsymbol{\mu}_x'\boldsymbol{\alpha}+\frac{A_3}{2}[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha}+\boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}]\tag{11}$$

If the distribution of x, f_1, and f_2 are unknown, given an α value, we can estimated the purity by

$$\hat{\text{purity}}(\boldsymbol{\alpha})=\hat{A}_0+\frac{\hat{A}_1}{2}+\hat{A}_2\hat{\boldsymbol{\mu}}_x'\boldsymbol{\alpha}+\frac{\hat{A}_3}{2}[\boldsymbol{\alpha}'\hat{\boldsymbol{\Sigma}}_x\boldsymbol{\alpha}+\boldsymbol{\alpha}'\hat{\boldsymbol{\mu}}_x\hat{\boldsymbol{\mu}}_x'\boldsymbol{\alpha}]\tag{12}$$

1.2.2 Optimization of $\boldsymbol{\alpha}$

Above Equation (11) has given us the purity funciton based on $\boldsymbol{\alpha}$. As well as the restriction of $\boldsymbol{\alpha}$ that $||\boldsymbol{\alpha}|| = 1$, we can use the method of Lagrange multiplier to find the solution of $\boldsymbol{\alpha}$ to maximize the data purity.

Besides, we also notice that, if standardization of baseline covariates is performed, the

- $\mu_x = [0, ..]'_p$

The estimation of purity in equation can be simplified as is

$$\begin{aligned} \text{purity}(\boldsymbol{\alpha}) =& A_0 + \frac{A_1}{2} + A_2\boldsymbol{\mu}'_x\boldsymbol{\alpha} + \frac{A_3}{2}[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}'_x\boldsymbol{\alpha}] \\ =& A_0 + \frac{A_1}{2} + \frac{A_3}{2}\alpha'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} \end{aligned} \tag{13}$$

Then the function to be optimzized with restriction $||\boldsymbol{\alpha}|| - 1 = 0$ can be defined as $h(\boldsymbol{\alpha}; \lambda)$:

$$\begin{aligned} h(\boldsymbol{\alpha}; \lambda) =& A_0 + \frac{A_1}{2} + \frac{A_3}{2}\alpha'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} + \lambda(\boldsymbol{\alpha}'\boldsymbol{\alpha} - 1) \\ =& A_0 + \frac{A_1}{2} - \lambda + \boldsymbol{\alpha}'(\frac{A_3}{2}\boldsymbol{\Sigma}_x + \lambda\boldsymbol{I})\boldsymbol{\alpha} \end{aligned} \tag{14}$$

When A_0, A_1, A_3 are known (constant), to maximize purity$(\boldsymbol{\alpha}) = \alpha'\boldsymbol{\Sigma}_x\boldsymbol{\alpha}$ subjects to $g(\boldsymbol{\alpha}) = 0$,. So by Lagrange multiplier, there is λ so that

$$\nabla\text{purity} = \lambda\nabla g$$

Note $\nabla g(\boldsymbol{\alpha}) = 2\boldsymbol{\alpha}$. On the other hand, $\nabla\text{purity}(\boldsymbol{\alpha}) = 2\boldsymbol{\Sigma}_x\boldsymbol{\alpha}$ as $\boldsymbol{\Sigma}_x$ is symmetirc. Thus we have $\boldsymbol{\Sigma}_x\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha}$. Therefore, $\boldsymbol{\alpha}$ is an eigenvector of $\boldsymbol{\Sigma}_x$ and λ is an eigenvalue of $\boldsymbol{\Sigma}_x$. The eigenvector of $\boldsymbol{\Sigma}_x$ with the largest eigenvalue can maximize the purity function.

1.2.3 Algorithm

Given the formulas of data purity and the solution of $\boldsymbol{\alpha}$, the algortihm to find the $\boldsymbol{\alpha}$ that maximize the purity as well as the max purity can be summerzied as:

1) Set an initial $\boldsymbol{\alpha}^{(0)}$ value. And fit the LME model.

$$\boldsymbol{Y} = \boldsymbol{S}(\boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\Gamma}(\boldsymbol{\alpha}'\boldsymbol{x})) + \boldsymbol{\epsilon}. \tag{15}$$

2) Estimate $\hat{\boldsymbol{\Sigma}}_x$ and get the $\hat{\lambda}^{(1)}$ and $\hat{\boldsymbol{\alpha}}^{(1)}$, which is the largest eigenvalue of $\hat{\boldsymbol{\Sigma}}_x$ and its corresponding eigenvector.

3) Estimate $\hat{\boldsymbol{\beta}}_1^{(1)}, \hat{\boldsymbol{\beta}}_2^{(1)}, \hat{\boldsymbol{\Gamma}}_1^{(1)}, \hat{\boldsymbol{\Gamma}}_2^{(1)}, \hat{\boldsymbol{D}_1}^{(1)}, \hat{D_2}^{(1)}$.

4) Plug in the above estimated values in equation 13 to get the estimated purity.

5) Wrap the 1-4 steps into a function and optimize the function with Newton Raphson method.

1.2.4 others

If we cannot assume $\boldsymbol{\mu}_x = 0$, then

we can use the method of Lagrange multiplier to find the solution of $\boldsymbol{\alpha}$ to maximize the data purity:

$$\begin{aligned} h(\boldsymbol{\alpha};\lambda) =& A_0 + \frac{A_1}{2} + A_2\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{A_3}{2}[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}] + \lambda(\boldsymbol{\alpha}'\boldsymbol{\alpha} - 1) \\ =& A_0 + \frac{A_1}{2} - \lambda + A_2\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{A_3}{2}\boldsymbol{\alpha}'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \frac{2\lambda}{A_3}\boldsymbol{I})\boldsymbol{\alpha} \end{aligned} \tag{16}$$

Based on the facts of matrix derivatives,

- $\frac{\partial AX}{\partial X} = A$
- $\frac{\partial X'AX}{\partial X} = X'(A + A')$

The first derivative of equation (13) is

$$\begin{aligned} \frac{\partial h(\boldsymbol{\alpha};\lambda)}{\partial \boldsymbol{\alpha}} =& A_2\boldsymbol{\mu}_x' + \frac{A_3}{2}\boldsymbol{\alpha}'[(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \frac{2\lambda}{A_3}\boldsymbol{I}) + (\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \frac{2\lambda}{A_3}\boldsymbol{I})'] \\ =& A_2\boldsymbol{\mu}_x' + A_3\boldsymbol{\alpha}'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \frac{2\lambda}{A_3}\boldsymbol{I}) \end{aligned} \tag{17}$$

Set Eq(14) = 0, we have

$$\hat{\boldsymbol{\alpha}} = -\frac{A_2}{A_3}(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \frac{2\lambda}{A_3}\boldsymbol{I})^{-1}\boldsymbol{\mu}_x \tag{18}$$

The second derivative of equation (13) is

$$A_3(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \frac{2\lambda}{A_3}\boldsymbol{I}) \tag{19}$$

The partial derivative of equation (13) w.r.t λ is:

$$\frac{\partial h(\boldsymbol{\alpha};\lambda)}{\partial \lambda} = \boldsymbol{\alpha}'\boldsymbol{\alpha} - 1 \tag{20}$$

Set Eq (17) = 0 and plug in the estimated $\hat{\boldsymbol{\alpha}}$ value in, we have

$$\begin{aligned}
&\boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \lambda_2\boldsymbol{I})^{-1}(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \lambda_2\boldsymbol{I})^{-1}\boldsymbol{\mu}_x - \frac{A_3^2}{A_2^2} \\
=&\boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}\boldsymbol{\mu}_x - \frac{A_3^2}{A_2^2} + 2\lambda_2\boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}\boldsymbol{\mu}_x + \lambda_2^2\boldsymbol{\mu}_x'\boldsymbol{\mu}_x \\
=&B_0\lambda_2^2 + 2B_1\lambda_2 + B_2 \\
=&(\sqrt{B_0}\lambda_2)^2 + 2\frac{B_1}{\sqrt{B_0}}\sqrt{B_0}\lambda_2 + \frac{B_1^2}{B_0} - \frac{B_1^2}{B_0} + B_2 \\
=&(\sqrt{B_0}\lambda_2 + \frac{B_1}{\sqrt{B_0}})^2 - (\frac{B_1^2}{B_0} - B_2) \\
=&0 \\
\rightarrow&\lambda_2 = \frac{1}{\sqrt{B_0}}(\sqrt{\frac{B_1^2}{\sqrt{B_0}} - B_2} - \frac{B_1}{\sqrt{B_0}}) = \sqrt{\frac{B_1^2}{B_0} - \frac{B_2}{\sqrt{B_0}}} - \frac{B_1}{B_0}
\end{aligned} \tag{21}$$

where

- $\lambda_2 = \frac{2\lambda}{A_3}$
- $B_0 = \boldsymbol{\mu}_x'\boldsymbol{\mu}_x$
- $B_1 = \boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}\boldsymbol{\mu}_x$
- $B_2 = \boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}\boldsymbol{\mu}_x - \frac{A_3^2}{A_2^2}$

Plug in the λ value, we could get the estimated $\boldsymbol{\alpha}$:

$$\hat{\boldsymbol{\alpha}} = -\frac{A_2}{A_3}(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x' + \lambda\boldsymbol{I})^{-1}\boldsymbol{\mu}_x$$

For example, if $\boldsymbol{X} \sim MVN(\mathbf{1}, I_{p_x})$, then

- $B_0 = \boldsymbol{\mu}_x'\boldsymbol{\mu}_x = p_x$
- $B_1 = \boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}\boldsymbol{\mu}_x = 0.5p_x$
- $B_2 = \boldsymbol{\mu}_x'(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}(\boldsymbol{\Sigma}_x + \boldsymbol{\mu}_x\boldsymbol{\mu}_x')^{-1}\boldsymbol{\mu}_x - \frac{A_3^2}{A_2^2} = 0.25p_x - \frac{A_3^2}{A_2^2}$

And

$$\lambda = \sqrt{\frac{B_1^2}{B_0} - \frac{B_2}{\sqrt{B_0}}} - \frac{B_1}{B_0} = \sqrt{0.25p_x - 0.25\sqrt{p_x} + \frac{A_3^2}{A_2^2}} - 0.5$$

where p_x is the dimension of $\boldsymbol{X}$