Elucidating gene networks in resistance diffuse large B-cell lymphoma

Will Patterson

8/28/2020

Jargon: An introduction

- **Canonical variate**: linear combinations of variables
- **Canonical variate pair**: two canonical variates (one from each set) with some non-zero correlation
- **Canonical correlation**: the correlation between canonical variate pairs

Canonical Correlation Analysis: Definition

Canonical Correlation Analysis (CCA) finds the relationship between two sets of variables by finding the maximally correlated linear combinations of variables.

Canonical Correlation Analysis: Nitty Gritty

Given: two sets of observations on same n observations, $\mathbf{X}_1$ and $\mathbf{X}_2$ of dimensions $n \times p_1$ and $n \times p_2$, standardized to mean zero and SD of one

Find: Weights $\mathbf{w}_1 \in \mathbb{R}^{p_1}$ and $\mathbf{w}_2 \in \mathbb{R}^{p_2}$ the objective function below is maximized

CCA objective function

$$\text{maximize}_{w_1,w_2} \mathbf{w}_1^T \mathbf{X}_1^T \mathbf{X}_2 \mathbf{w}_2$$

$$\text{subject to } ||\mathbf{w}_1||^2 \leq 1, ||\mathbf{w}_2||^2 \leq 1, P_1(\mathbf{w}_1) \leq c_1, P_2(\mathbf{w}_2) \leq c_2$$

(where P_1 and P_2 are penalty functions, usually lasso, L_1)

Sparse **Multiple** CCA

Sparse multiple CCA (or sparse mCCA) is an extensible form of sparse CCA, generalizable for any K data sets $\mathbf{X}_1, ..., \mathbf{X}_K$, where data set k contains p_k features. The objective function then takes the form:

Sparse mCCA objective function

$$\text{maximize}_{\mathbf{w}_1,...,\mathbf{w}_K} \sum_{i<j} \mathbf{w}_i^T \mathbf{X}_i^T \mathbf{X}_j \mathbf{w}_j \text{ subject to } ||\mathbf{w}_i||^2 \leq 1, P_i(\mathbf{w}_i) \leq c_i, \forall i$$

Extension of sparse mCCA to binary outcomes

Witten and Tibshirani (2009) suggest an extension of sparse mCCA that allows for the incorporation of a two-class outcome. Their method simply treats this $\mathbb{R}^{n \times 1}$ matrix as a third data set. Their objective function takes the form:

Sparse mCCA objective function with binary variables

$$\text{maximize}_{\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3} \mathbf{w}_1^T \mathbf{X}_1^T \mathbf{X}_2 \mathbf{w}_2 + \mathbf{w}_1^T \mathbf{X}_1^T \mathbf{y} \mathbf{w}_3 + \mathbf{w}_2^T \mathbf{X}_2^T \mathbf{y} \mathbf{w}_3$$

$$\text{subject to } ||\mathbf{w}_i||^2 \leq 1, P_i(\mathbf{w}_i) \leq c_i, \forall i$$