New Fuzzy Methods For Feature Selection

Contents

1. Abstract

2. Objective Functions

From the book named SPEC, we sort the eigenvalues of S increasingly and we let ξ_1 be the first corresponding eigenvector.

In the equation, $x_{t,i}$ denotes the value of instance $\mathbf{x}_t$ on feature $\mathbf{f}_i$. $NH(\mathbf{x})$ and $NM(\mathbf{x})$ denote the nearest points to $\mathbf{x}$ in the data with the same and different labels, respectively, and $\|\cdot\|$ is a distance measurement.

$$\mathbf{S}_{ij}^{FIS} = \begin{cases} \frac{1}{n_l}, & y_i = y_j = l \\ 0, & otherwise \end{cases} \quad \max_{\mathbf{F}_{sub}} \sum_{F \in \mathbf{F}_{sub}} \hat{\mathbf{f}}^\top \hat{\mathbf{S}} \hat{\mathbf{f}}, \quad \mathbf{S}_{i,j}^{REL} = \begin{cases} 1 & i = j \\ -\frac{1}{k} & x_j \in NH(\mathbf{x}_i) \\ \frac{1}{(c-1)k} & x_j \in NM(\mathbf{x}_i, CL(\mathbf{x}_i)) \end{cases}$$

TABLE 4.1: The similarity matrices and feature vectors used in different algorithms.

Algorithm	Sample Similarity Matrix	Feature Normalization
Fisher Score	$\mathbf{S}^{FIS}$	$\tilde{\mathbf{f}} = \frac{\mathbf{D}^{\frac{1}{2}}\mathbf{f}}{\|\mathbf{D}^{\frac{1}{2}}\mathbf{f}\|}$, $\hat{\mathbf{f}} = \frac{\tilde{\mathbf{f}} - \tilde{\mathbf{f}}^\top \xi_1 \xi_1}{\sqrt{1 - (\tilde{\mathbf{f}}^\top \xi_1)^2}}$
ReliefF	$\mathbf{S}^{REL}$	$\hat{\mathbf{f}} = \frac{\mathbf{D}^{\frac{1}{2}}\mathbf{f}}{\|\mathbf{D}^{\frac{1}{2}}\mathbf{f}\|}$
Trace Ratio Criterion	$\mathbf{L}_b - \lambda^* \mathbf{L}_w$	$\hat{\mathbf{f}} = \mathbf{f}$

From the article named SLEP, ...

Table 3: Applicability of the SLEP package

Penalty	Problem	Function	Description	Section
Lasso	$\min_{\mathbf{x}:\|\mathbf{x}\|_1\le z} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2$	eplb	Euclidean projection onto the ℓ_1 ball	2.1
	$\min_{\mathbf{x}} f(\mathbf{x})+\lambda\|\mathbf{x}\|_1$	LeastR	Least squares loss	2.2
		LogisticR	Logistic loss	2.3
	$\min_{\mathbf{x}:\|\mathbf{x}\|_1\le z} f(\mathbf{x})$	LeastC	Least squares loss	2.4
		LogisticC	Logistic loss	2.5
	$\min_{\mathbf{x}\ge\mathbf{0}} f(\mathbf{x})+\lambda\|\mathbf{x}\|_1$	nnLeastR	Least squares loss	2.6
		nnLogisticR	Logistic loss	2.7
	$\min_{\mathbf{x}:\|\mathbf{x}\|_1\le z,\mathbf{x}\ge\mathbf{0}} f(\mathbf{x})$	nnLeastC	Least squares loss	2.8
		nnLogisticC	Logistic loss	2.9
group Lasso[1]	$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\|\mathbf{x}\|_q$	epp	ℓ_q-regularized Euclidean projection	3.1
	$\min_{\mathbf{x}} f(\mathbf{x})+\lambda\|\mathbf{x}\|_{q,1}$	glLeastR	Group Lasso	3.2
		glLogisticR		3.3
		mtLeastR	Multi-task learning	3.4
		mtLogisticR		3.5
		mcLeastR	Multi-class/task learning	3.6
		mcLogisticR		3.7
	$\min_{\mathbf{x}:\|\mathbf{x}\|_{2,1}\le z} f(\mathbf{x})$	mtLeastC	Multi-task learning	4.1
		mtLogisticC		4.2
		mcLeastC	Multi-class/task learning	4.3
		mcLogisticC		4.4
fused Lasso	$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda_1\|\mathbf{x}\|_1+\lambda_2\sum_{i=1}^{p-1}\|x_i-x_{i+1}\|$	flsa	fused Lasso signal approximator	5.1
	$\min_{\mathbf{x}} f(\mathbf{x})+\lambda_1\|\mathbf{x}\|_1+\lambda_2\sum_{i=1}^{p-1}\|x_i-x_{i+1}\|$	fusedLeastR	Least Squares Loss	5.2
		fusedLogisticR	Logistic Loss	5.3
sparse inverse covariance	$\max_{\Theta\succ 0} \log\|\Theta\|-\langle S,\Theta\rangle-\lambda\|\Theta\|_1$	spaInvCov	sparse inverse covariance estimation	6
sparse group Lasso[2]	$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda_1\|\mathbf{x}\|_1+\lambda_2\sum_{i=1}^{g} w_i\|\mathbf{x}_{G_i}\|_2$	altra	Moreau-Yosida Regularization	7.1
	$\min_{\mathbf{x}} f(\mathbf{x})+\lambda_1\|\mathbf{x}\|_1+\lambda_2\sum_{i=1}^{g} w_i\|\mathbf{x}_{G_i}\|_2$	sgLeastR	Least Squares Loss	7.2
		sgLogisticR	Logistic Loss	7.3
		mc_sgLeastR	Logistic Loss	7.4
tree structured group Lasso[3]	$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\sum_{i=1}^{g} w_i\|\mathbf{x}_{G_i}\|_2$	altra	Moreau-Yosida Regularization	8.1
		general_altra		
	$\min_{\mathbf{x}} f(\mathbf{x})+\lambda\sum_{i=1}^{g} w_i\|\mathbf{x}_{G_i}\|_2$	tree_LeastR	Least Squares Loss	8.2
		tree_LogisticR	Logistic Loss	8.3
		tree_mcLeastR	Least Squares Loss	8.4
		tree_mcLogisticR	Logistic Loss	8.5
		tree_mcLeastR	Least Squares Loss	8.6
		tree_mcLogisticR	Logistic Loss	8.7
overlapping group Lasso[4]	$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2+\lambda\sum_{i=1}^{g} w_i\|\mathbf{x}_{G_i}\|_2$	overlapping	Moreau-Yosida Regularization	9.1
	$\min_{\mathbf{x}} f(\mathbf{x})+\lambda\sum_{i=1}^{g} w_i\|\mathbf{x}_{G_i}\|_2$	overlapping_LeastR	Least Squares Loss	9.2
		overlapping_LogisticR	Logistic Loss	9.3
Ordered Tree	$\min_{\mathbf{x}\in P} \frac{1}{2}\|\mathbf{x}-\mathbf{v}\|_2^2$	orderTree	Euclidean Projection onto P	10.1
	$\min_{\mathbf{x}\in P} f(\mathbf{x})+\lambda\|\mathbf{x}\|_1$	orderLeastC	Least Squares Loss	10.2
Least Squares Loss		pathSolutionLeast	Pathwise solutions	12
Logistic Loss		pathSolutionLogistic		
trace norm	$\min_W \frac{1}{2}\|\|XW-Y\|\|_F^2+\lambda\|\|W\|\|_*$	accel_grad_mlr	Linear regression	13.1
		mat_primal	Linear regression	13.4
		mat_dual	Linear regression	13.5
	$\min_W \frac{1}{2}\sum_{i=1}^{k}\|\|X_i w_i-Y_i\|\|_2^2+\lambda\|\|W\|\|_*$	accel_grad_mtl	Multi-task learning	13.2
	$\min_W \sum_{i=1}^{n}\ell(y_i,\mathrm{Tr}(W^TX_i))+\lambda\|\|W\|\|_*$	accel_grad_mc	Matrix classification	13.3

[1]: The ℓ_1/ℓ_q-norm is defined as the summation of ℓ_q-norm of the non-overlapping groups.
[2]: In the sparse group Lasso, the indices G_i do not overlap, i.e. $G_i\cap G_j=\emptyset,\forall i\ne j$.
[3]: In the tree structured group Lasso, the indices G_i overlap. However, note that, G_i's follow the tree structure, as depicted in Figure 11.
[4]: In the overlapping group Lasso, the indices G_i may overlap, without the restriction in the tree structured group Lasso.

3. Comparative Map — Weights(x_v^h, y^h)

With Pearson's method, if each variance of x_v and y are unitary, then $W_1^3 = \dfrac{\overline{x_1 y} - \overline{x_1} \cdot \overline{y}}{\sum_{v=1}^n (\overline{x_v y} - \overline{x_v} \cdot \overline{y})}$.

If each mean of x_v and y are all zero, then $W_1^3 = \dfrac{\overline{x_1 y}}{\sum_{v=1}^n \overline{x_v y}} = \dfrac{\sum x_1^h y^h}{\sum x_1^h y^h + \cdots + \sum x_n^h y^h}$.

$$\frac{\partial W}{\partial y^i} = u' v^{-1} + u(-v^{-2}) v' = \frac{x_1^i y^i}{\sum x_1^h y^h + \cdots + \sum x_n^h y^h} - \frac{\sum x_1^h y^h}{[\sum x_1^h y^h + \cdots + \sum x_n^h y^h]^2} (x_1^i + \cdots + x_n^i).$$

$$j \geq 2 \Rightarrow \frac{\partial W}{\partial x_j^i} = c(-v^{-2}) v' = -\frac{\sum x_1^h y^h}{[\sum x_1^h y^h + \cdots + \sum x_n^h y^h]^2} \cdot \cancel{x_j^i} \cdot y^i.$$

$$\frac{\partial W}{\partial x_1^i} = \frac{y^i}{\sum x_1^h y^h + \cdots + \sum x_n^h y^h} - \frac{\sum x_1^h y^h}{[\sum x_1^h y^h + \cdots + \sum x_n^h y^h]^2} \cdot y^i.$$

By Taylor's expansion, and by the combinatorics $\underbrace{AAACCE \sim \dfrac{5!}{3!0!2!0!1!}}_{c_1 \text{ counts how many indexes } = A?\ 3.}$ permutations,

$$f(x_1, \cdots, x_n) = f(x^0) + \sum_{g=1}^{\infty} \frac{1}{\cancel{g!}} \sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_g}^{n} \frac{\cancel{g!}}{c_1! \cdots c_g!} \prod_{k=1}^{g} (x_{i_k} - x_{i_k}^0) \frac{\partial^g f(x^0)}{\partial x_{i_1} \cdots \partial x_{i_g}}. \tag{1}$$

With YNFN-FS, we should numerically get all the first derivatives $\dfrac{\partial f(x^0)}{\partial x_i}$ at the middle point $x^0 = \dfrac{x^i + x^f}{2}$. Then $\dfrac{\partial^2 f(x^0)}{\partial x_{i_1} \partial x_{i_2}}$ and successively, until a Taylor's good approach (no one with MAPE above 20%); but here we need, whenever $n_{FP} = 3$, at least 2^n series (n-cubic regions).

4. Conclusions and Future Work

May/17th/2023 Release* by Vinicius Claudino Ferraz.

*Out of charity, there is no salvation at all.