Learning structured data via flow represented interactions of SVM *Indirect Kernel Learning*

S. Szedmak, E. Parrado-Hernández, J. Shawe-Taylor, H. Meng and J.D.R. Farquhar

{ss03v,eph,jst}@ecs.soton.ac.uk.

ISIS Group
School of Electronics and Computer Science
University of Southampton, UK

March 2005

Outline

- Beyond binary classification problems

Outline

- Beyond binary classification problems
- Common framework from optimisation point of view

Outline

- Beyond binary classification problems
- Common framework from optimisation point of view
- One optimisation problem: flow interaction

Outline

- Beyond binary classification problems
- Common framework from optimisation point of view
- One optimisation problem: flow interaction
- Efficient solution to the global problem

Outline

- Beyond binary classification problems
- Common framework from optimisation point of view
- One optimisation problem: flow interaction
- Efficient solution to the global problem
- Numerical results

Outline

- Beyond binary classification problems
- Common framework from optimisation point of view
- One optimisation problem: flow interaction
- Efficient solution to the global problem
- Numerical results
- Fruitful discussion!

Multi-stuff classification problems

Things to do with binary classifiers

- Multifeature, multiview, cotraining (Meng), (Dasgupta 2001)
- Multilabel
- Multiclass
- Multikernel
- Multitask

Combine different representation vectors belonging to the same object. Labels are shared among representations.

Multi-stuff classification problems

Things to do with binary classifiers

- Multifeature, multiview, cotraining
- Multilabel (Taskar 2003), (Bartlett 2004), (Rousu)
- Multiclass
- Multikernel
- Multitask

Labels have structure. Hierarchy of labels, non-exclusive labels

Multi-stuff classification problems

Things to do with binary classifiers

- Multifeature, multiview, cotraining
- Multilabel
- Multiclass (Rifkin 2004)
- Multikernel
- Multitask

Label can be one out of $M > 2$. Special attention to very large number of classes where perhaps there is a chance for error correcting codes.

Multi-stuff classification problems

Things to do with binary classifiers

- Multifeature, multiview, cotraining
- Multilabel
- Multiclass
- Multikernel (Lanckriet 2004), (Bach 2004)
- Multitask

Kernel is a combination of kernel matrices. Same optimisation problem to determine the classifier and the kernel.

Multi-stuff classification problems

Things to do with binary classifiers

- Multifeature, multiview, cotraining
- Multilabel
- Multiclass
- Multikernel
- Multitask (Micchelli 2004)

General framework for multilabel, multiview and multikernel

Kernel synthesis - Learning Kernels

Philosophy of our advanced approach

- **Common practise:** Choosing best kernels and/or features

Kernel synthesis - Learning Kernels

Philosophy of our advanced approach

- **Common practise:** Choosing best kernels and/or features
- **Here:** Combining them into an ensemble which outperforms the components

Kernel synthesis - Learning Kernels

Philosophy of our advanced approach

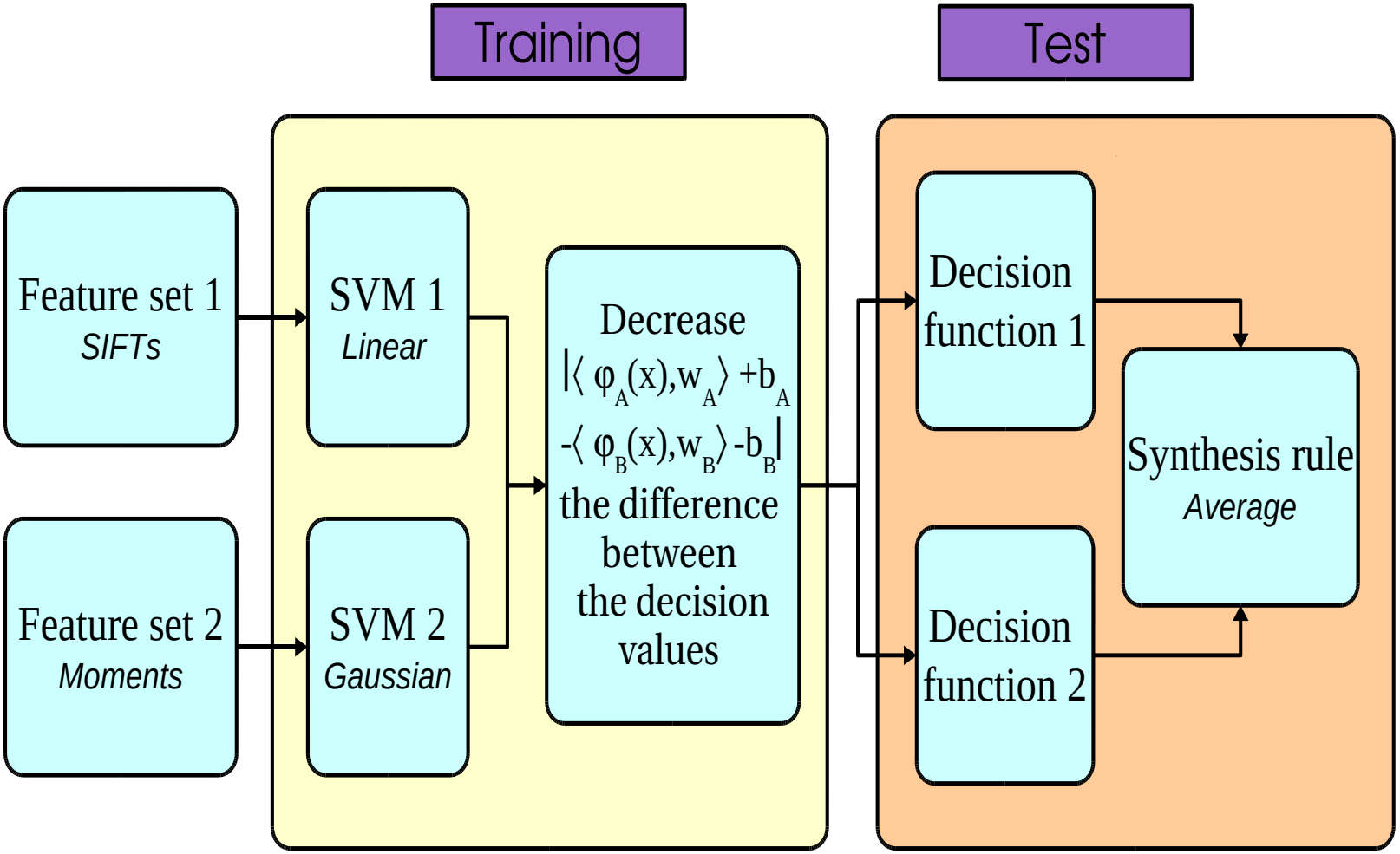
- **Common practise:** Choosing best kernels and/or features
- **Here:** Combining them into an ensemble which outperforms the components
- **Boosting idea:** The weak learners are connected by a simple convex combination

Kernel synthesis - Learning Kernels

Philosophy of our advanced approach

- **Common practise:** Choosing best kernels and/or features
- **Here:** Combining them into an ensemble which outperforms the components
- **Boosting idea:** The weak learners are connected by a simple convex combination
- **Our Extension** weak learners connected through a multivariate nonlinear function.

Combination via features



Nomenclature

- Data: abstract objects $\Omega = \{\omega_i\}_{i=1}^{m}$

Nomenclature

- Data: abstract objects $\Omega = \{\omega_i\}_{i=1}^m$
- K different representations:
  $S_k = \{\boldsymbol{s}_i^{(k)} = (\boldsymbol{x}_i^{(k)}, \boldsymbol{y}_i^{(k)})\}_{i=1}^m, k = 1, \ldots, K$
  - Feature vectors: $\boldsymbol{x}_i^{(k)} \in \mathcal{X}_k$
  - Label vectors: $\boldsymbol{y}_i^{(k)} \in \{-1, +1\}^K, Y_k = \text{diag}(y_i^k)$

Nomenclature

- Data: abstract objects $\Omega = \{\omega_i\}_{i=1}^m$
- K different representations:
  $S_k = \{\boldsymbol{s}_i^{(k)} = (\boldsymbol{x}_i^{(k)}, \boldsymbol{y}_i^{(k)})\}_{i=1}^m, k = 1, \ldots, K$
  - Feature vectors: $\boldsymbol{x}_i^{(k)} \in \mathcal{X}_k$
  - Label vectors: $\boldsymbol{y}_i^{(k)} \in \{-1, +1\}^K, Y_k = \text{diag}(y_i^k)$
- Mappings: $\boldsymbol{\phi}_k : \mathcal{X}_k \rightarrow \Re^{n_k}$

More nomenclature

- Binary classifiers: $h_k : \Re^{n_k} \rightarrow \Re$,
  $h_k(\boldsymbol{x}^{(k)}) = \boldsymbol{\phi}_k(\boldsymbol{x}^{(k)})^T \boldsymbol{w}_k + b_k$

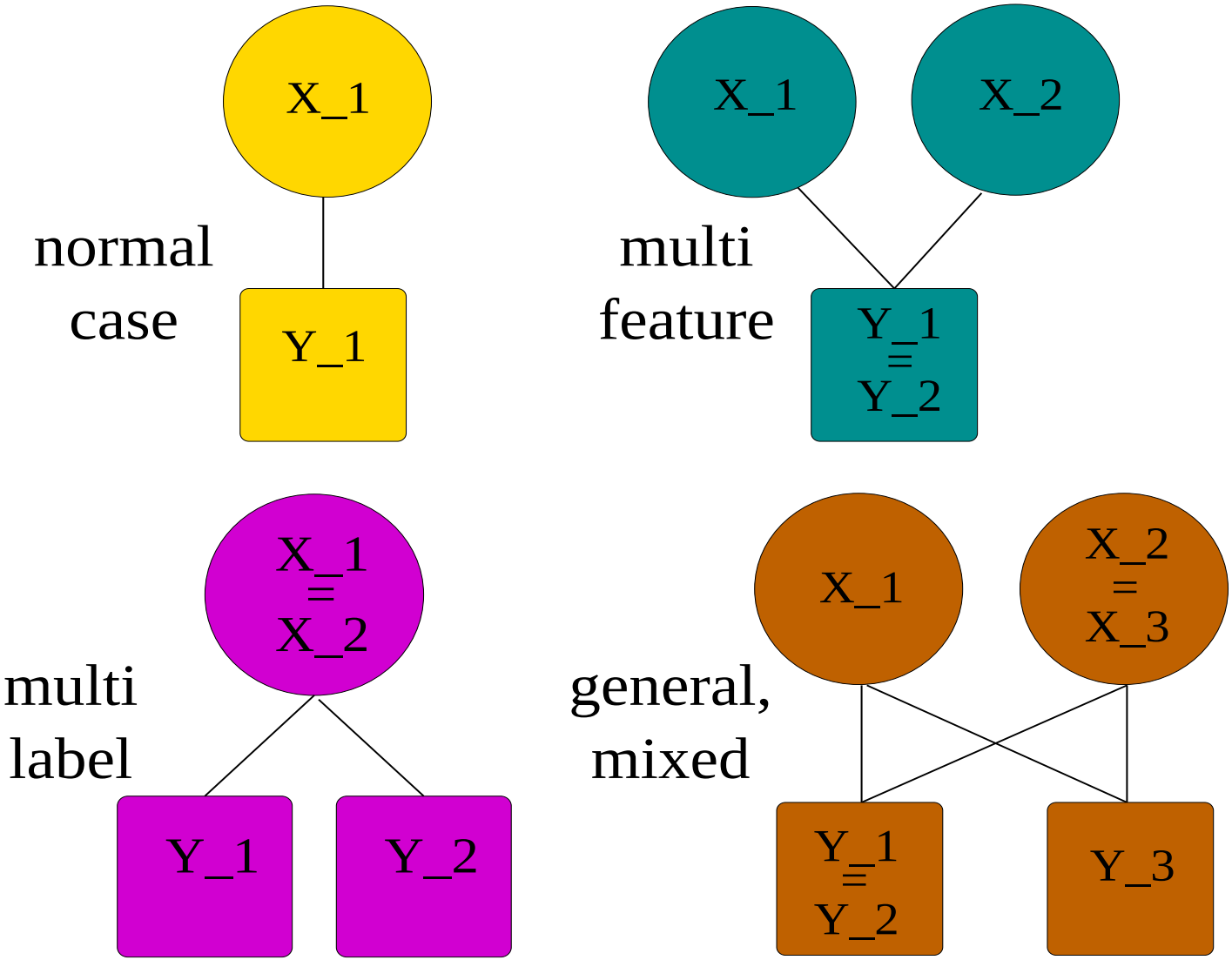
More nomenclature

- Binary classifiers: $h_k : \Re^{n_k} \longrightarrow \Re$,
  $h_k(\boldsymbol{x}^{(k)}) = \boldsymbol{\phi}_k(\boldsymbol{x}^{(k)})^T \boldsymbol{w}_k + b_k$
- Overall classifier: $\boldsymbol{f} : \Re^{n_k} \longrightarrow \{-1, +1\}^K$,
  $(\hat{\boldsymbol{y}}_i^{(k)})_j = (\boldsymbol{f})_j(h_1, \ldots, h_{n_k}), j = 1, \ldots, K$

More nomenclature

- Binary classifiers: $h_k : \Re^{n_k} \longrightarrow \Re$,
  $h_k(\boldsymbol{x}^{(k)}) = \boldsymbol{\phi}_k(\boldsymbol{x}^{(k)})^T \boldsymbol{w}_k + b_k$
- Overall classifier: $\boldsymbol{f} : \Re^{n_k} \longrightarrow \{-1, +1\}^K$,
  $(\hat{\boldsymbol{y}}_i^{(k)})_j = (\boldsymbol{f})_j(h_1, \ldots, h_{n_k})$, $j = 1, \ldots, K$
- Constant Design Matrix $G \in \{-1, 0, +1\}^{L \times K}$

The expressable cases



Primal problem

$$\min \frac{1}{2}\sum_{k}^{K} w_k^T w_k + \sum_{k}^{K} C_k \mathbf{1}^T \xi_k + \sum_{l}^{L} D_l \mathbf{1}^T (\eta_l^+ + \eta_l^-)$$

with respect to $\{w_j, b_j, \xi_j\}_{j=1}^{K}, \quad \{\eta_i^+, \eta_i^-\}_{i=1}^{L}$

---

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k \ ,$$

$$\sum_{k}^{K} g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- , \ l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^{K} \geq 0 \ , \quad \{\eta_l^+, \ \eta_l^-\}_{l=1}^{L} \geq 0 \ .$$

Primal problem

$$\min \boxed{\frac{1}{2}\sum_{k}^{K} w_k^T w_k} + \sum_{k}^{K} C_k \mathbf{1}^T \xi_k + \sum_{l}^{L} D_l \mathbf{1}^T (\eta_l^+ + \eta_l^-)$$

with respect to $\{w_j, b_j, \ \xi_j \ \}_{j=1}^{K}, \ \ \{\eta_i^+, \eta_i^-\}_{i=1}^{L}$

---

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k \ ,$$

$$\sum_{k}^{K} g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- , \ l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^{K} \geq 0 \ , \ \ \{\eta_l^+, \ \eta_l^-\}_{l=1}^{L} \geq 0 \ .$$

Primal problem

$$\min \frac{1}{2}\sum_{k}^{K} w_k^T w_k + \boxed{\sum_{k}^{K} C_k \mathbf{1}^T \xi_k} + \sum_{l}^{L} D_l \mathbf{1}^T (\eta_l^+ + \eta_l^-)$$

with respect to $\{w_j, b_j, \boxed{\xi_j}\}_{j=1}^{K}, \quad \{\eta_i^+, \eta_i^-\}_{i=1}^{L}$

---

subject to

$$\boxed{Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k},$$

$$\sum_{k}^{K} g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^-, \; l = 1, \ldots, L$$

$$\boxed{\{\xi_k\}_{k=1}^{K} \geq 0}, \quad \{\eta_l^+, \; \eta_l^-\}_{l=1}^{L} \geq 0 \; .$$

Primal problem

$$\min \frac{1}{2}\sum_{k}^{K} w_k^T w_k + \sum_{k}^{K} C_k \mathbf{1}^T \xi_k + \boxed{\sum_{l}^{L} D_l \mathbf{1}^T (\eta_l^+ + \eta_l^-)}$$

with respect to $\{w_j, b_j, \xi_j\}_{j=1}^{K}$, $\boxed{\{\eta_i^+, \eta_i^-\}_{i=1}^{L}}$

---

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k \ ,$$

$$\boxed{\sum_{k}^{K} g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- , \ l = 1, \ldots, L}$$

$$\{\xi_k\}_{k=1}^{K} \geq 0 \ , \ \boxed{\{\eta_l^+, \ \eta_l^-\}_{l=1}^{L} \geq 0}.$$

Dual problem

$$\min \ \frac{1}{2}\sum_{k}^{K} u_k^T \mathcal{K}_k u_k - \sum_{k}^{K} \mathbf{1}\alpha_k$$

with respect to

$$\{\alpha_k \ , \ u_k \ \}_{k=1}^{K}, \ \{\beta_l\}_{l=1}^{L} \ , \ u_k = Y_k \alpha_k - \sum_{l}^{L} g_{lk}\beta_l$$

subject to

$$\mathbf{1}^T u_k = 0 \ ,$$

$$0 \leq \alpha_k \leq C_k \ , \ \ -D_l \leq \beta_l \leq D_l \ ,$$

$$k = 1, \ldots, K, \ l = 1, \ldots, L.$$

Dual problem

$$\min \boxed{\frac{1}{2}\sum_{k}^{K} u_k^T \mathcal{K}_k u_k - \sum_{k}^{K} \mathbf{1}\alpha_k}$$

with respect to

$$\boxed{\{\alpha_k}, \; u_k \}_{k=1}^{K}, \; \{\beta_l\}_{l=1}^{L} \; , \; u_k = Y_k \alpha_k - \sum_{l}^{L} g_{lk}\beta_l$$

subject to

$$\mathbf{1}^T u_k = 0 \; ,$$

$$\boxed{0 \leq \alpha_k \leq C_k}, \quad -D_l \leq \beta_l \leq D_l \; ,$$

$$k = 1, \ldots, K, \; l = 1, \ldots, L.$$

Dual problem

$$\min \ \frac{1}{2}\sum_{k}^{K} u_k^T \mathcal{K}_k u_k - \sum_{k}^{K} \mathbf{1}\alpha_k$$

with respect to

$$\{\alpha_k \ , \boxed{u_k}\}_{k=1}^{K}, \boxed{\{\beta_l\}_{l=1}^{L}}, \boxed{u_k = Y_k\alpha_k - \sum_{l}^{L} g_{lk}\beta_l}$$

subject to

$$\boxed{\mathbf{1}^T u_k = 0},$$

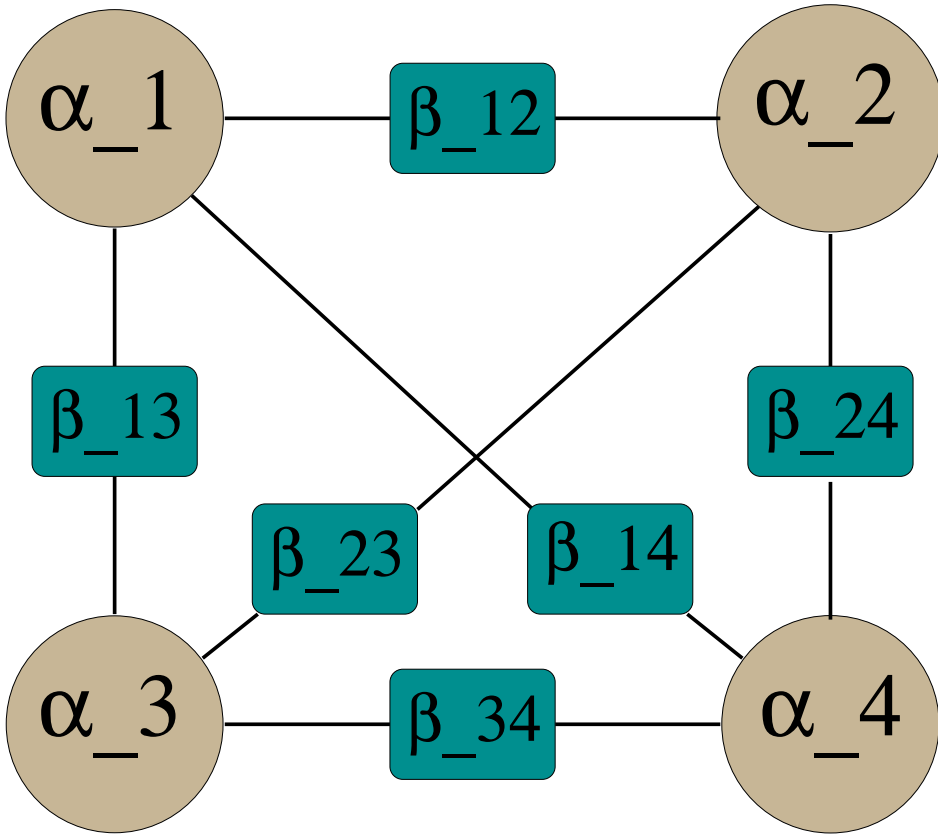
$$0 \leq \alpha_k \leq C_k \ , \ \boxed{-D_l \leq \beta_l \leq D_l},$$

$$k = 1, \ldots, K, \ l = 1, \ldots, L.$$

Interaction among nodes

$$G = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

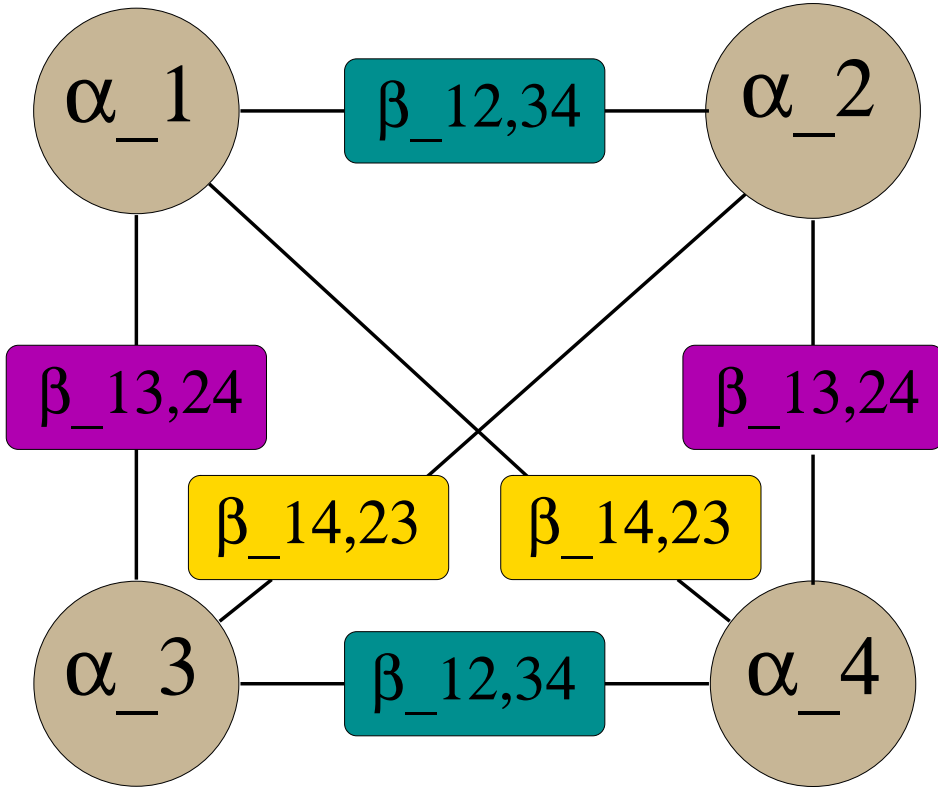
$$\{h_1\} \Leftrightarrow \{h_2\} \quad \{h_1\} \Leftrightarrow \{h_3\}$$
$$\{h_1\} \Leftrightarrow \{h_4\} \quad \{h_2\} \Leftrightarrow \{h_3\}$$
$$\{h_2\} \Leftrightarrow \{h_4\} \quad \{h_3\} \Leftrightarrow \{h_4\}$$



Interaction among subsets of nodes

$$G = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

$$\{h_1, h_3\} \Leftrightarrow \{h_2, h_4\}$$
$$\{h_1, h_2\} \Leftrightarrow \{h_3, h_4\}$$
$$\{h_1, h_4\} \Leftrightarrow \{h_2, h_3\}$$



Synthetised Kernel Matrix Q

	α_1	$\cdots$	α_K	β_1	$\cdots$	β_L
α_1	$Y_1\mathcal{K}_1Y_1$	$\cdots$	$\emptyset$	$-g_{11}Y_1\mathcal{K}_1$	$\cdots$	$-g_{1L}Y_1\mathcal{K}_1$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
α_K	$\emptyset$	$\cdots$	$Y_K\mathcal{K}_KY_K$	$-g_{K1}Y_K\mathcal{K}_K$	$\cdots$	$-g_{KL}Y_K\mathcal{K}_K$
β_1	$-g_{11}\mathcal{K}_1Y_1$	$\cdots$	$-g_{1K}\mathcal{K}_KY_K$	$\sum_k g_{1k}\mathcal{K}_kg_{k1}$	$\cdots$	$\sum_k g_{1k}\mathcal{K}_kg_{kL}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
β_L	$-g_{L1}\mathcal{K}_1Y_1$	$\cdots$	$-g_{LK}\mathcal{K}_KY_K$	$\sum_k g_{Lk}\mathcal{K}_kg_{k1}$	$\cdots$	$\sum_k g_{Lk}\mathcal{K}_kg_{kL}$

Synthetised Kernel Matrix Q

	α_1	$\cdots$	α_K	β_1	$\cdots$	β_L
α_1	$\boxed{Y_1\mathcal{K}_1Y_1}$	$\cdots$	$\emptyset$	$-g_{11}Y_1\mathcal{K}_1$	$\cdots$	$-g_{1L}Y_1\mathcal{K}_1$
$\vdots$	$\vdots$	$\boxed{\ddots}$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
α_K	$\emptyset$	$\cdots$	$\boxed{Y_K\mathcal{K}_KY_K}$	$-g_{K1}Y_K\mathcal{K}_K$	$\cdots$	$-g_{KL}Y_K\mathcal{K}_K$
β_1	$-g_{11}\mathcal{K}_1Y_1$	$\cdots$	$-g_{1K}\mathcal{K}_KY_K$	$\sum_k g_{1k}\mathcal{K}_kg_{k1}$	$\cdots$	$\sum_k g_{1k}\mathcal{K}_kg_{kL}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
β_L	$-g_{L1}\mathcal{K}_1Y_1$	$\cdots$	$-g_{LK}\mathcal{K}_KY_K$	$\sum_k g_{Lk}\mathcal{K}_kg_{k1}$	$\cdots$	$\sum_k g_{Lk}\mathcal{K}_kg_{kL}$

Component of the Gradient w.r.t Q

Gradient with respect to Q: $(q_1, \ldots, q_{K+L})^T = Qz$,

where $z = (\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_L)$

is expressable using the substitution $u_k = Y_k \alpha_k - \sum_l^L g_{kl} \beta_l$ by

$$\boxed{\begin{aligned} q_k &= \mathcal{K}_k u_k, \; k = 1, \ldots, K, \\ q_{K+l} &= \textstyle\sum_k^K g_{kl} Y_k q_k, \; l = 1, \ldots, L, \end{aligned}}$$

which reduces the complexity:

$O((K+L)^2 m^2) \Rightarrow O(K m^2)$

The augmentation matrix allows similar simplifi cation.

Efficient Nested Implementation

To exploit the special structure of the problem.

Efficient Nested Implementation

To exploit the special structure of the problem.

- **Augmented Lagrangian Method:** To eliminate equality constrains

Efficient Nested Implementation

To exploit the special structure of the problem.

- **Augmented Lagrangian Method:** To eliminate equality constrains
  - **Conditional Gradient Method:** To solve problem with fixed Lagrangian multipliers

Efficient Nested Implementation

To exploit the special structure of the problem.

- **Augmented Lagrangian Method:** To eliminate equality constrains
  - **Conditional Gradient Method:** To solve problem with fixed Lagrangian multipliers
    - **Linear Programming problem:** To derive the next approximation of the optimum

Efficient Nested Implementation

To exploit the special structure of the problem.

- **Augmented Lagrangian Method:** To eliminate equality constrains
  - **Conditional Gradient Method:**
    To solve problem with fixed Lagrangian multipliers
    - **Linear Programming problem:**
      To derive the next approximation of the optimum
      - This problem is a trivial one with linear complexity.
        It gives the key element of the efficiency!

A way of relaxation

$$\min \frac{1}{2}\sum_k^K w_k^T w_k + \sum_k^K C_k \mathbf{1}^T \xi_k + \sum_l^L D_l \mathbf{1}^T(\eta_l^+ + \eta_l^-)$$

w. r. t. $\{w_k, b_k, \xi_k\}_{k=1}^K,\ \{\eta_i^+, \eta_i^-\}_{i=1}^L$

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k \ ,$$

$$\sum_k^K g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- \ ,$$

$$l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^K \geq 0\ \{\eta_l^+,\ \eta_l^-\}_{l=1}^L \geq 0.$$

$$\min \frac{1}{2}w_k^T w_k + (C_k + \hat{D})\mathbf{1}^T \xi_k + \hat{D}\mathbf{1}^T \tau_k \ ,$$

with respect to

$$\{w_k, b_k, \xi_k, \tau_k\}_{k=1}^K$$

subject to

$$Y_k(\phi_k(X)w_k + b_k) = 1 - \xi_k + \tau_k \ ,$$

$$\xi_k \geq 0,\ \tau_k \geq 0,\ k = 1, \ldots, K,$$

A way of relaxation

$$\min \frac{1}{2}\sum_k^K w_k^T w_k + \sum_k^K C_k \mathbf{1}^T \xi_k + \sum_l^L D_l \mathbf{1}^T(\eta_l^+ + \eta_l^-)$$

w. r. t. $\{w_k, b_k, \xi_k\}_{k=1}^K,\ \{\eta_i^+, \eta_i^-\}_{i=1}^L$

subject to

$$Y_k(\phi_k(X)w_k + b_k) \boxed{\geq 1 - \xi_k},$$

$$\sum_k^K g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- \ ,$$

$$l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^K \geq 0\ \{\eta_l^+,\ \eta_l^-\}_{l=1}^L \geq 0.$$

$$\min \frac{1}{2}w_k^T w_k + (C_k + \hat{D})\mathbf{1}^T \xi_k + \hat{D}\mathbf{1}^T \tau_k \ ,$$

with respect to

$$\{w_k, b_k, \xi_k, \tau_k\}_{k=1}^K$$

subject to

$$Y_k(\phi_k(X)w_k + b_k) \boxed{= 1 - \xi_k + \tau_k},$$

$$\xi_k \geq 0,\ \tau_k \geq 0,\ k = 1, \ldots, K,$$

A way of relaxation

$$\min \frac{1}{2}\sum_k^K w_k^T w_k + \boxed{\sum_k^K C_k \mathbf{1}^T \xi_k} + \sum_l^L D_l \mathbf{1}^T(\eta_l^+ + \eta_l^-)$$

w. r. t. $\{w_k, b_k, \xi_k\}_{k=1}^K,\ \{\eta_i^+, \eta_i^-\}_{i=1}^L$

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k \ ,$$

$$\sum_k^K g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- \ ,$$

$$l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^K \geq 0\ \{\eta_l^+,\ \eta_l^-\}_{l=1}^L \geq 0.$$

$$\min \frac{1}{2}w_k^T w_k + \boxed{(C_k + \hat{D})\mathbf{1}^T \xi_k} + \hat{D}\mathbf{1}^T \tau_k \ ,$$

with respect to

$$\{w_k, b_k, \xi_k, \tau_k\}_{k=1}^K$$

subject to

$$Y_k(\phi_k(X)w_k + b_k) = 1 - \xi_k + \tau_k \ ,$$

$$\xi_k \geq 0,\ \tau_k \geq 0,\ k = 1, \ldots, K,$$

A way of relaxation

$$\min \frac{1}{2}\sum_k^K w_k^T w_k + \sum_k^K C_k \mathbf{1}^T \xi_k + \sum_l^L D_l \mathbf{1}^T(\eta_l^+ + \eta_l^-)$$

w. r. t. $\{w_k, b_k, \xi_k\}_{k=1}^K,\ \{\eta_i^+, \eta_i^-\}_{i=1}^L$

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k\ ,$$

$$\boxed{\sum_k^K g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^-},$$

$$l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^K \geq 0\ \{\eta_l^+,\ \eta_l^-\}_{l=1}^L \geq 0.$$

$$\min \frac{1}{2}w_k^T w_k + (C_k + \hat{D})\mathbf{1}^T \xi_k + \hat{D}\mathbf{1}^T \tau_k\ ,$$

with respect to

$$\{w_k, b_k, \xi_k, \tau_k\}_{k=1}^K$$

subject to

$$Y_k(\phi_k(X)w_k + b_k) = 1 - \xi_k + \tau_k\ ,$$

$$\xi_k \geq 0,\ \tau_k \geq 0,\ k = 1, \ldots, K,$$

A way of relaxation

$$\min \quad \frac{1}{2}\sum_k^K w_k^T w_k + \sum_k^K C_k \mathbf{1}^T \xi_k + \boxed{\sum_l^L D_l \mathbf{1}^T (\eta_l^+ + \eta_l^-)}$$

w. r. t. $\{w_k, b_k, \xi_k\}_{k=1}^K, \ \{\eta_i^+, \eta_i^-\}_{i=1}^L$

subject to

$$Y_k(\phi_k(X)w_k + b_k) \geq 1 - \xi_k \ ,$$

$$\sum_k^K g_{lk}(\phi_k(X)w_k + b_k) = \eta_l^+ - \eta_l^- \ ,$$

$$l = 1, \ldots, L$$

$$\{\xi_k\}_{k=1}^K \geq 0 \ \{\eta_l^+, \ \eta_l^-\}_{l=1}^L \geq 0.$$

$$\min \quad \frac{1}{2} w_k^T w_k + \boxed{(C_k + \hat{D})\mathbf{1}^T \xi_k} + \boxed{\hat{D}\mathbf{1}^T \tau_k},$$

with respect to

$$\{w_k, b_k, \xi_k, \tau_k\}_{k=1}^K$$

subject to

$$Y_k(\phi_k(X)w_k + b_k) = 1 - \xi_k + \tau_k \ ,$$

$$\xi_k \geq 0, \ \tau_k \geq 0, \ k = 1, \ldots, K,$$

Dual of the relaxed problem

$$\min \tfrac{1}{2}\alpha_k^T \mathcal{K}_k \alpha_k - \mathbf{1}\alpha_k$$

with respect to α_k,

subject to

$$\mathbf{1}^T Y \alpha_k = 0,$$

$$-\hat{D} \leq \alpha_k \leq C_k,$$

Dual of the relaxed problem

$$\min \tfrac{1}{2}\alpha_k^T \mathcal{K}_k \alpha_k - \mathbf{1}\alpha_k$$

with respect to α_k,

subject to

$$\mathbf{1}^T Y \alpha_k = 0,$$

$$\boxed{-\hat{D}} \leq \alpha_k \leq C_k,$$

$-\hat{D} = 0$ in the regular SVM

Improved computation time

Computation times for each dataset using different optimization methods

	Optimization method	
Data sets	Matlab Optimization Toolbox	Augmented Lagrangian
Bike	1320s	8.5s
Persons	1080s	10.8s
Motorbikes	$>$ 1 day(43h)	43s
Airplanes	$>$ 1 day	42s
Faces	$>$ 1 day	65s

LAVA dataset bike and persons
Caltech dataset motorbikes and airplanes

XEROX 7 class dataset

Methods	Acc (STD)
Baseline:kMeans(1K) SVM	85.0( )
GMM(250) SVM_2k PCA(40) on SIFT+Moment	88.4(2.8)
GMM(250) SVM_2k_V2 PCA(40) on SIFT+Moment	90.4(2.5)
kMeans(1K) SVM_2K on SIFT+Moment	91.3(2.5)
kMeans(1K) SVM_2K_V2 on SIFT+Moment	92.9(1.5)

not published

Discussion

Possible explanation to the good results:

- Optimisation problem that we actually solve is not as ill-possed as the original
- Against normal SVM
  - Maybe we are facing data distributions without higher order moments
  - Data in high-dim space enables to align every class in a hyperplane and draw a separating hyperplane.