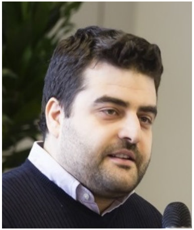
An Epigraphic Splitting Technique for Sparse Multiclass SVM

Jean-Christophe PESQUET

Laboratoire d'Informatique Gaspard Monge - CNRS
Univ. Paris-Est, France

SPARS 2015

In collaboration with

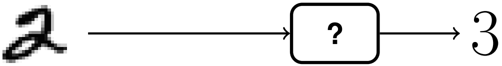


G. Chierchia
Telecom ParisTech

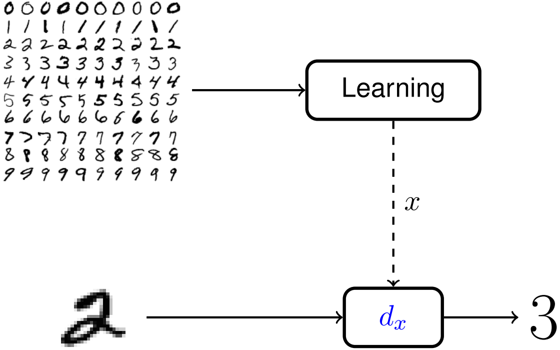
N. Pustelnik
ENS Lyon

B. Pesquet-Popescu
Telecom ParisTech

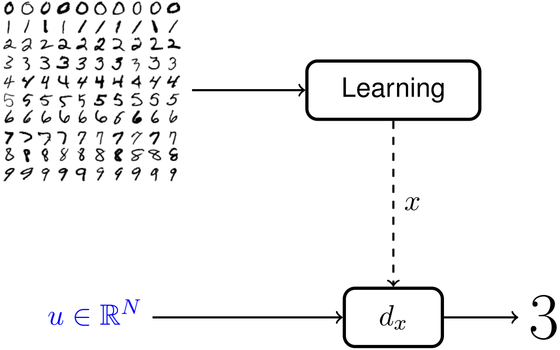
Supervised classification



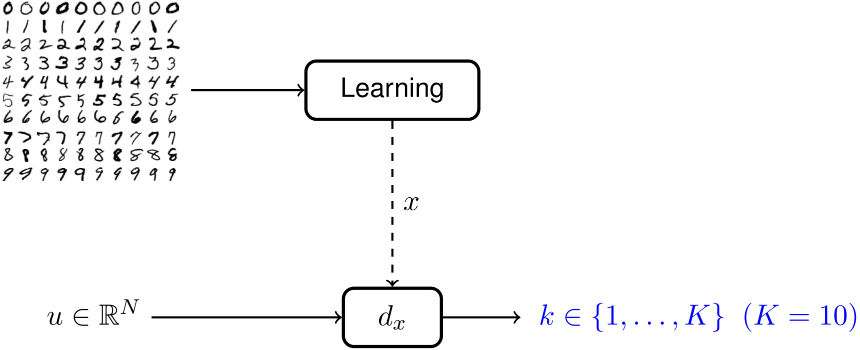
Supervised classification



Supervised classification



Supervised classification



Supervised classification

$$\begin{tikzcd}
& \{(u_\ell, z_\ell) \mid \ell = 1, \dots, L\} \arrow[r] & \boxed{\text{Learning}} \arrow[d, dashed, "x"] & \\
u \in \mathbb{R}^N \arrow[rr] & & \boxed{d_x} \arrow[r] & k \in \{1, \dots, K\} \quad (K = 10)
\end{tikzcd}$$

Notation

- **Training set** of size L for K classes:

$$\mathcal{S} = \big\{(u_\ell, z_\ell) \in \mathbb{R}^N \times \{1,\dots,K\} \;\big|\; \ell \in \{1,\dots,L\}\big\}$$

<u>examples:</u> $u_\ell =$  and $z_\ell = 2$

$u_\ell =$  and $z_\ell = 9$

Notation

- **Training set** of size L for K classes:

$$\mathcal{S} = \big\{(u_\ell, z_\ell) \in \mathbb{R}^N \times \{1, \ldots, K\} \;\big|\; \ell \in \{1, \ldots, L\}\big\}$$

<u>examples:</u> $u_\ell =$  and $z_\ell = 2$

$u_\ell =$  and $z_\ell = 9$

- $\phi\colon \mathbb{R}^N \to \mathbb{R}^M$: mapping from the input space onto an M-dimensional feature space

$\Rightarrow$ **linearization**

<u>examples:</u> convolution networks [Mirowski et al. 2008]

scattering coefficients [Brunat, Mallat 2013]

Objective

- The predictor relies on K discriminating functions $D_k \colon \mathbb{R}^N \to \mathbb{R}$, $k \in \{1, \ldots, K\}$:

$$\boxed{D_k(u) = \phi(u)^\top x^{(k)}}$$

where $x^{(k)} \in \mathbb{R}^M$.

- The classifier selects the class that best matches an observation:

$$d_x(u) = \arg \max_{1 \leqslant k \leqslant K} D_k(u)$$

- Objective of the learning stage: estimate $x = (x^{(1)}, \ldots, x^{(K)})$ to correctly predict the input-output pair (u_ℓ, z_ℓ) for every $\ell \in \{1, \ldots, L\}$.

Multi-class SVM

- Objective of learning :

$$z_\ell = \underset{1 \leqslant k \leqslant K}{\arg\max}\; \phi(u_\ell)^\top x^{(k)}$$

Multi-class SVM

- Objective of learning :

$$\boxed{z_\ell = \arg\max_{1 \leqslant k \leqslant K} \phi(u_\ell)^\top x^{(k)}}$$

$$\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) < 0$$

Multi-class SVM

- Objective of learning :

$$\boxed{z_\ell = \arg\max_{1 \leqslant k \leqslant K} \phi(u_\ell)^\top x^{(k)}}$$

$$\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) < 0$$

[relax the strict inequality] $\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant -\mu_\ell$

with $\mu_\ell > 0$

Multi-class SVM

- Objective of learning :

$$\boxed{z_\ell = \arg\max_{1 \leqslant k \leqslant K} \phi(u_\ell)^\top x^{(k)}}$$

$$\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) < 0$$

[relax the strict inequality] $\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant -\mu_\ell$

[deal with unfeasible constraints] $\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant \zeta^{(\ell)} - \mu_\ell$

with $\mu_\ell > 0$ and $\zeta^{(\ell)} \geqslant 0$.

Multi-class SVM

- Objective of learning :

$$\boxed{z_\ell = \underset{1\leqslant k\leqslant K}{\arg\max}\ \phi(u_\ell)^\top x^{(k)}}$$

$$\Leftrightarrow \quad \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) < 0$$

[relax the strict inequality] $\Leftrightarrow \quad \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant -\mu_\ell$

[deal with unfeasible constraints] $\Leftrightarrow \quad \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant \zeta^{(\ell)} - \mu_\ell$

with $\mu_\ell > 0$ and $\zeta^{(\ell)} \geqslant 0$.

- **Convex optimization problem** [Crammer-Singer 2001]

$$\underset{(x,\zeta)}{\text{minimize}} \qquad \text{s. t.}$$

$$(\forall \ell \in \{1,\dots,L\}) \quad \begin{cases} \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) + \mu_\ell \leqslant \zeta^{(\ell)} \\ \zeta^{(\ell)} \geqslant 0 \end{cases}$$

Multi-class SVM

- Objective of learning :

$$\boxed{z_\ell = \arg\max_{1\leqslant k\leqslant K} \phi(u_\ell)^\top x^{(k)}}$$

$$\Leftrightarrow \quad \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) < 0$$

[relax the strict inequality] $\Leftrightarrow \quad \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant -\mu_\ell$

[deal with unfeasible constraints] $\Leftrightarrow \quad \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant \zeta^{(\ell)} - \mu_\ell$

with $\mu_\ell > 0$ and $\zeta^{(\ell)} \geqslant 0$.

- **Convex optimization problem** [Crammer-Singer 2001]

  $\rightarrow$ maximize distance of samples to separating hyperplanes

$$\underset{(x,\zeta)}{\text{minimize}} \quad \frac{1}{2}\|x\|_2^2 \qquad \text{s. t.}$$

$$(\forall \ell \in \{1,\dots,L\}) \quad \begin{cases} \max_{k\neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) + \mu_\ell \leqslant \zeta^{(\ell)} \\ \zeta^{(\ell)} \geqslant 0 \end{cases}$$

Multi-class SVM

- Objective of learning :

$$z_\ell = \arg\max_{1 \leqslant k \leqslant K} \phi(u_\ell)^\top x^{(k)}$$

$$\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) < 0$$

[relax the strict inequality] $\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant -\mu_\ell$

[deal with unfeasible constraints] $\Leftrightarrow \quad \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \leqslant \zeta^{(\ell)} - \mu_\ell$

with $\mu_\ell > 0$ and $\zeta^{(\ell)} \geqslant 0$.

- **Convex optimization problem** [Crammer-Singer 2001]
  - $\rightarrow$ maximize distance of samples to separating hyperplanes
  - $\rightarrow$ minimize margin violation

$$\underset{(x,\zeta)}{\text{minimize}} \quad \frac{1}{2}\|x\|_2^2 + \lambda \sum_{\ell=1}^{L} \zeta^{(\ell)} \quad \text{s. t.}$$

$$(\forall \ell \in \{1, \ldots, L\}) \quad \begin{cases} \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) + \mu_\ell \leqslant \zeta^{(\ell)} \\ \zeta^{(\ell)} \geqslant 0 \end{cases}$$

where $\lambda \in]0, +\infty[$.

Standard SVM

$$\underset{(x,\zeta)}{\text{minimize}} \ \frac{1}{2}\|x\|_2^2 + \lambda \sum_{\ell=1}^{L} \zeta^{(\ell)} \quad \text{s. t.}$$

$$(\forall \ell \in \{1,\ldots,L\}) \qquad \begin{cases} \max\limits_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) + \mu_\ell \leqslant \zeta^{(\ell)} \\ \zeta^{(\ell)} \geqslant 0 \end{cases}$$

- Solution via Lagrangian duality
  [Crammer-Singer, 2001,Tsochantaridis 2005, Joachims 2009, ...]

- **Preventing overfitting with sparsity priors**
  [Quattoni et al 2009, Jalali et al. 2010, Bach et al. 2012]

Proposed formulation

$$\underset{(x,\zeta)\in(\mathbb{R}^{(M+1)})^K\times\mathbb{R}^L}{\text{minimize}} \quad g(x)+\lambda\sum_{\ell=1}^{L}\zeta^{(\ell)} \quad \text{s.t.}$$

$$(\forall \ell\in\{1,\dots,L\})\quad h_\ell(T_\ell x)\leqslant\zeta^{(\ell)}$$

or equivalently

$$\underset{x\in(\mathbb{R}^{(M+1)})^K}{\text{minimize}}\ g(x)+\lambda\sum_{\ell=1}^{L}h_\ell(T_\ell x)$$

with $\begin{cases} h_\ell(y^{(\ell)}) = \max\limits_{1\leqslant k\leqslant K} y^{(\ell,k)} + r_\ell^{(k)} \\ T_\ell\, x = \left[\,\phi(u_\ell)^\top(x^{(k)}-x^{(z_\ell)})\,\right]_{1\leqslant k\leqslant K}. \end{cases}$ with $r_\ell^{(k)} = \begin{cases} 0, & \text{if } k=z_\ell, \\ \mu_\ell, & \text{otherwise,} \end{cases}$

Proposed formulation

Regularized formulation:

$$\underset{x\in(\mathbb{R}^{(M+1)})^K}{\text{minimize}} \; g(x) + \lambda \sum_{\ell=1}^{L} h_\ell(T_\ell x)$$

Constrained formulation:

$$\underset{x\in(\mathbb{R}^{(M+1)})^K}{\text{minimize}} \; g(x) \quad \text{s. t.} \quad \sum_{\ell=1}^{L} h_\ell(T_\ell x) \leqslant \eta \in \,]0,+\infty[$$

with $\begin{cases} h_\ell(y^{(\ell)}) = \max\limits_{1\leqslant k\leqslant K} y^{(\ell,k)} + r_\ell^{(k)} \\ T_\ell\, x = \left[\, \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)}) \,\right]_{1\leqslant k\leqslant K}. \end{cases}$ with $r_\ell^{(k)} = \begin{cases} 0, & \text{if } k = z_\ell, \\ \mu_\ell, & \text{otherwise,} \end{cases}$

Proposed formulation

$$\underset{x\in(\mathbb{R}^{(M+1)})^K}{\text{minimize}} \quad g(x) + \lambda \sum_{\ell=1}^{L} h_\ell(T_\ell x)$$

Constrained formulation:

$$\underset{(x,\zeta)\in(\mathbb{R}^{(M+1)})^K\times\mathbb{R}^L}{\text{minimize}} \quad g(x) \quad \text{s. t.} \quad \sum_{\ell=1}^{L} \zeta^{(\ell)} \leqslant \eta$$
$$(\forall \ell \in \{1,\dots,L\}) \quad h_\ell(T_\ell x) \leqslant \zeta^{(\ell)}$$

with $\begin{cases} h_\ell(y^{(\ell)}) = \max\limits_{1\leqslant k\leqslant K} y^{(\ell,k)} + r_\ell^{(k)} \\ T_\ell\, x = \left[\,\phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\,\right]_{1\leqslant k\leqslant K}. \end{cases}$ with $r_\ell^{(k)} = \begin{cases} 0, & \text{if } k = z_\ell, \\ \mu_\ell, & \text{otherwise,} \end{cases}$

- **Contributions: exact resolution of sparse SVM**
  - $\rightarrow$ regularized formulation using proximal splitting
  - $\rightarrow$ constrained formulation using epigraphical proximal splitting

Algorithmic solution

$$\widehat{y} \in \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}} \sum_{r=1}^{R} g_r(A_r y) \quad \text{s.t.} \quad \begin{cases} B_1 y \in C_1, \\ \vdots \\ B_S y \in C_S, \end{cases}$$

- $(\forall r \in \{1,\ldots,R\})$ $g_r\colon \mathbb{R}^{N_r} \to]-\infty,+\infty]$ convex, lsc, and proper,
- $(\forall r \in \{1,\ldots,R\})$ $A_r \in \mathbb{R}^{N_r \times N}$,
- $(\forall s \in \{1,\ldots,S\})$ $B_s \in \mathbb{R}^{M_s \times N}$,
- $(\forall s \in \{1,\ldots,S\})$ C_s is a nonempty closed convex subset of $\mathbb{R}^{M_s}$.

Algorithmic solution

$$\widehat{y} \in \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}} \sum_{r=1}^{R} g_r(A_r y) \quad \text{s.t.} \quad \begin{cases} \boxed{B_1 y \in C_1}, \\ \vdots \\ \boxed{B_S y \in C_S}, \end{cases}$$

- $(\forall r \in \{1, \dots, R\})$ $g_r \colon \mathbb{R}^{N_r} \to]-\infty, +\infty]$ convex, lsc, and proper,
- $(\forall r \in \{1, \dots, R\})$ $A_r \in \mathbb{R}^{N_r \times N}$,
- $(\forall s \in \{1, \dots, S\})$ $B_s \in \mathbb{R}^{M_s \times N}$,
- $(\forall s \in \{1, \dots, S\})$ C_s is a nonempty closed convex subset of $\mathbb{R}^{M_s}$.

$\rightsquigarrow$ Primal-dual proximal splitting algorithms

Algorithmic solution

For $n = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
y^{[n]} = \sum_{r=1}^{R} \omega_r u_r^{[n]} + \sum_{s=1}^{S} \omega_s \overline{u}_s^{[n]} \\
\text{For } r = 1, \ldots, R \\
\quad \left\lfloor \begin{array}{l} w_{1,r}^{[n]} = u_r^{[n]} - \gamma_\ell A_r^\top v_r^{[n]} \\ w_{2,r}^{[n]} = v_r^{[n]} + \gamma_\ell A_r u_r^{[n]} \end{array} \right. \\
\text{For } s = 1, \ldots, S \\
\quad \left\lfloor \begin{array}{l} \overline{w}_{1,s}^{[n]} = \overline{u}_s^{[n]} - \gamma_n B_s^\top \overline{v}_s^{[n]} \\ \overline{w}_{2,s}^{[n]} = \overline{u}_s^{[n]} + \gamma_n B_s \overline{u}_s^{[n]} \end{array} \right. \\
p_1^{[n]} = \sum_{r=1}^{R} \omega_r w_{1,r}^{[n]} + \sum_{s=1}^{S} \omega_s \overline{w}_{1,s}^{[n]} \\
\text{For } r = 1, \ldots, R \\
\quad \left\lfloor \begin{array}{l} \boxed{p_{2,r}^{[n]} = w_{2,r}^{[n]} - \dfrac{\gamma_n}{\omega_r} \operatorname{prox}_{\frac{\omega_r}{\gamma_n} g_r} \left(\dfrac{\omega_r}{\gamma_n} w_{2,r}^{[n]} \right)} \\ q_{1,r}^{[n]} = p_1^{[n]} - \gamma_n (A_r^\top p_{2,r}^{[n]}) \\ q_{2,r}^{[n]} = p_{2,r}^{[n]} + \gamma_n (A_r p_1^{[n]}) \\ \text{Update } u_r^{[n+1]} \text{ and } v_r^{[n+1]} \end{array} \right. \\
\text{For } s = 1, \ldots, S \\
\quad \left\lfloor \begin{array}{l} \boxed{\overline{p}_{2,s}^{[n]} = \overline{w}_{2,r}^{[n]} - \dfrac{\gamma_n}{\omega_s} P_{C_s} \left(\dfrac{\omega_s}{\gamma_n} \overline{w}_{2,r}^{[n]} \right)} \\ \overline{q}_{1,s}^{[n]} = \overline{p}_1^{[n]} - \gamma_n (B_s^\top \overline{p}_{2,s}^{[n]}) \\ \overline{q}_{2,s}^{[n]} = \overline{p}_{2,s}^{[n]} + \gamma_n (B_s \overline{p}_1^{[n]}) \\ \text{Update } \overline{u}_s^{[n+1]} \text{ and } \overline{v}_s^{[n+1]} \end{array} \right.
\end{array}
\right.
$$

$\longleftarrow$ Under technical assumptions, $(y^{[n]})_{n\in\mathbb{N}}$ generated by M+SFBF [Combettes-Briceño-Arias, 2011] converges to $\hat{x}$

$\longleftarrow$ Proximity operator computation

$\longleftarrow$ Projection computation

Proximity operator

Compute $\operatorname{prox}_{\lambda h}$ where $h(u) = \max_{1\leqslant k\leqslant K} u^{(k)} + r^{(k)}$.

- Solution:

$$(\forall u \in \mathbb{R}^K) \quad \operatorname{prox}_{\lambda h}(u) = u - P_{S_\lambda}(u+r)$$

with

$$S_\lambda = \{u = (u^{(k)})_{1\leqslant k\leqslant K} \in [0,+\infty[^K \mid \sum_{k=1}^{K} u^{(k)} = \lambda\}.$$

Epigraphical projection

Compute P_C where $C = \{(u,\zeta) \in \mathbb{R}^K \times \mathbb{R} \,|\, h(u) \leqslant \zeta\}$.

- Epigraphical projection $P_C = P_{\operatorname{epi} h}$

Theorem

Let $h \in \Gamma_0(\mathbb{R}^K)$ be such that $\operatorname{dom} h$ is open. The projector onto the epigraph of h is given by

$$\big(\forall (u,\zeta) \in \mathbb{R}^K \times \mathbb{R}\big) \qquad P_{\operatorname{epi} h}(u,\zeta) = (p,\theta)$$

where

$$\begin{cases} p &= \operatorname{prox}_{\frac{1}{2}(\max\{h-\zeta,0\})^2}(u), \\ \theta &= \max\{h(p),\zeta\}. \end{cases}$$

Epigraphical projection

Compute P_C where $C = \{(u,\zeta) \in \mathbb{R}^K \times \mathbb{R} \,|\, h(u) \leqslant \zeta\}$.

- Epigraphical projection $P_C = P_{\operatorname{epi} h}$ with $h(u) = \max\limits_{1 \leqslant k \leqslant K} u^{(k)} + r^{(k)}$
- Solution: Let $\left(\nu^{(k)}\right)_{1 \leqslant k \leqslant K}$ be the sequence $\left(u^{(k)} + r^{(k)}\right)_{1 \leqslant k \leqslant K}$ sorted in ascending order, $\nu^{(0)} = -\infty$ and $\nu^{(K+1)} = +\infty$.

  Then, $P_{\operatorname{epi} h}(u,\zeta) = (p,\theta)$ with

$$\begin{cases} p = \Big(\min\{u^{(k)}, \theta - r^{(k)}\} \Big)_{1 \leqslant k \leqslant K} \\ \theta = \frac{1}{K - \overline{k} + 2} \left(\zeta + \sum_{k=\overline{k}}^{K} \nu^{(k)} \right), \end{cases}$$

  where $\overline{k}$ is the unique integer in $\{1, \ldots, K+1\}$ such that $\nu^{(\overline{k}-1)} < \theta \leqslant \nu^{(\overline{k})}$ with the convention $\sum_{k=K+1}^{K} \cdot = 0$.

Experimental setup

MNIST database

- $N = 28 \times 28$
- $K = 10$
- 10000 test images

Scattering network

- Feature mapping: $\phi \colon \mathbb{R}^{28\times 28} \mapsto \mathbb{R}^{14\times 14\times 81}$

Experiment 1

Classification errors with quadratic or sparse regularization

samples per class (L/K)	ℓ_2-SVM	$\ell_{1,\infty}$-SVM
3	27.06 %	**25.64** %
5	16.32 %	**13.59** %
10	11.00 %	**9.40** %
15	10.12 %	**7.68** %
20	7.78 %	**5.67** %
30	6.48 %	**5.46** %
50	4.22 %	**3.73** %
100	3.69 %	**3.13** %

Experiment 2

$$h(x) = \sum_{\ell=1}^{L} \max\left\{0, \mu_\ell + \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}$$

Comparison with other loss functions

Experiment 2

$$h(x) = \sum_{\ell=1}^{L} \max\left\{0, \mu_\ell + \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}$$

Comparison with other loss functions

- squared hinge loss [Blondel et al. 2013]

$$h(x) = \sum_{\ell=1}^{L} \sum_{k \neq z_\ell} \left(\max\left\{0, \mu_\ell + \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}\right)^2$$

Experiment 2

$$h(x) = \sum_{\ell=1}^{L} \max\left\{0, \mu_\ell + \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}$$

Comparison with other loss functions

- squared hinge loss [Blondel et al. 2013]
- logistic loss [Krishnapuram et al. 2005]

$$h(x) = \sum_{\ell=1}^{L} \log\left(1 + \sum_{k \neq z_\ell} \exp\left\{\mu_\ell + \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}\right)$$

Experiment 2

$$h(x) = \sum_{\ell=1}^{L} \max\left\{0, \mu_\ell + \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}$$

Comparison with other loss functions

- squared hinge loss [Blondel et al. 2013]
- logistic loss [Krishnapuram et al. 2005]
- one-vs-all [Laporte et al. 2014]
  binary SVM with $\widetilde{z}_\ell$ being equal to 1 if $z_\ell = k$, and -1 otherwise

$$h(x) = \sum_{\ell=1}^{L} \left(\max\left\{0, \mu_\ell + \widetilde{z}_\ell\, \phi(u_\ell)^\top x^{(k)}\right\}\right)^2$$

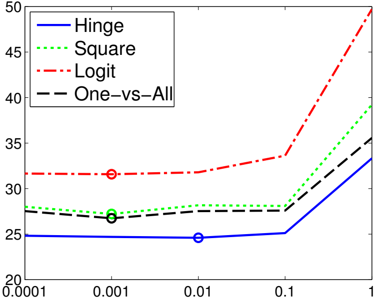
Experiment 2

$$h(x) = \sum_{\ell=1}^{L} \max\left\{0, \mu_\ell + \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}$$

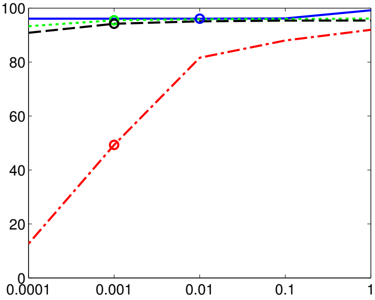
Comparison with other loss functions

Results with $L/K = 3$

% of classification errors vs λ^{-1}



% of zero coefficients vs λ^{-1}



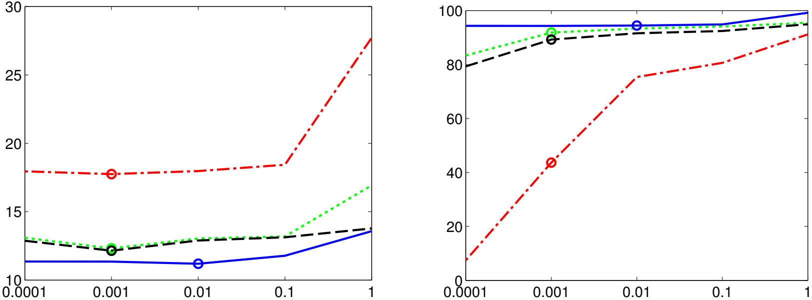
Experiment 2

$$h(x) = \sum_{\ell=1}^{L} \max\left\{0, \mu_\ell + \max_{k \neq z_\ell} \phi(u_\ell)^\top (x^{(k)} - x^{(z_\ell)})\right\}$$

Comparison with other loss functions

Results with $L/K = 10$

% of classification errors vs λ^{-1} — *% of zero coefficients vs* λ^{-1}



Experiment 3

Comparison of execution times

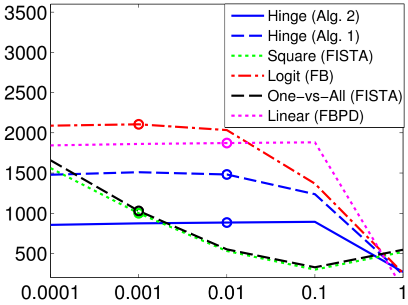
- hinge loss $\rightsquigarrow$ *proposed algorithms*
- hinge loss $\rightsquigarrow$ linear programming
- squared hinge loss $\rightsquigarrow$ FISTA
- one-vs-all $\rightsquigarrow$ FISTA

Experiment 3

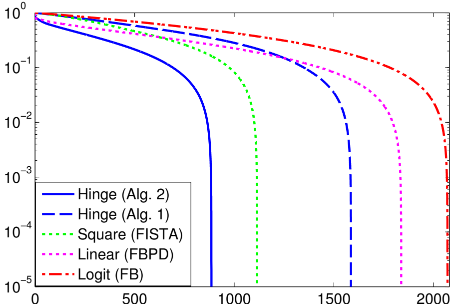
Comparison of execution times

Results with $L/K = 3$

Execution time vs λ^{-1}



distance to $x^{[+\infty]}$ vs time

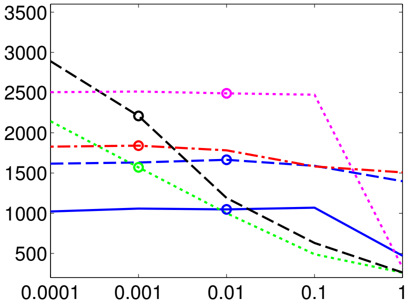


Experiment 3

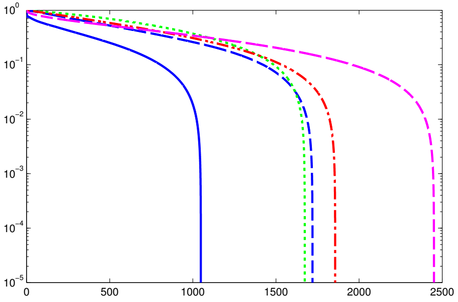
Comparison of execution times

Results with $L/K = 10$

Execution time vs λ^{-1}



distance to $x^{[+\infty]}$ *vs time*



Conclusion

- Revisit sparse SVM with epigraphical tools
- Computationally efficient algorithms
- Imposing sparsity improves results
- Other applications of epigraphical tools for sparse signal recovery: hyperspectral imagery, pulse shape design,...

Some references

G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu
A proximal approach for sparse multiclass SVM
2015, arXiv: 1501.03669.

G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu
Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools
to appear in *Signal, Image and Video Processing*, 2015.

P. L. Combettes and J.-C. Pesquet
Proximal splitting methods in signal processing
in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*,
H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz editors. Springer-Verlag, New York, pp. 185-212, 2011.

N. Komodakis and J.-C. Pesquet
Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems
to appear in *IEEE Signal Processing Magazine*, 2015.