Bilevel programming

D.A. Forsyth, with lots of stuff from Chris Fricke

Bilevel programming

What is Bilevel Programming?

- 'A mathematical program that contains an optimization problem in the constraints.' *
- Evolved in two ways:
  - A logical extension of mathematical programming.
  - Generalisation of a particular problem in game theory (Stackelberg Game).

*Bracken and McGill, *Operations Research*, Vol. 21 (1973)

Why we care:

This sounds like structure learning

Interesting SVM formulation

Maybe could do structure learning for continuous probs.

General Bilevel Programming Problem (Bard, 1998)

$$\min_{x \in X} F(x, y)$$

s.t.

$$G(x, y) \leq 0$$

$$\min_{y \in Y} f(x, y)$$

s.t.

$$g(x, y) \leq 0$$

$$x, y \geq 0$$

Neat SVM formulation

- Recall that when we solve an SVM, we need to
  - choose regularization parameter
  - perhaps, choose box for w
- Usual strategy
  - evaluate estimated error for various lambda with cross-validation
  - choose best
- i.e. solve optimization problem (in lambda)
  - which depends on inner optimization problems (SVM's in folds)

Neat SVM formulation

$$\begin{aligned}
&\underset{\mathbf{W},\mathbf{b},\lambda,\overline{\mathbf{w}}}{\text{minimize}} && \Theta(\mathbf{W},\mathbf{b}) \\
&\text{subject to} && \lambda_{\mathsf{lb}} \le \lambda \le \lambda_{\mathsf{ub}}, \quad \overline{\mathbf{w}}_{\mathsf{lb}} \le \overline{\mathbf{w}} \le \overline{\mathbf{w}}_{\mathsf{ub}}, \\
& && \text{and for } t = 1,\ldots,T,
\end{aligned}$$

$$(\mathbf{w}^t, b_t) \in \underset{\substack{-\overline{\mathbf{w}} \le \mathbf{w} \le \overline{\mathbf{w}} \\ b \in \mathbb{R}}}{\arg\min} \left\{ \frac{\lambda}{2} \| \mathbf{w} \|_2^2 + \sum_{j \in \overline{\mathcal{N}}_t} \max\left(1 - y_j(\mathbf{x}_j'\mathbf{w} - b), 0\right) \right\}.$$

w^t, b^t are the folds, one per fold

W is a matrix of these

wbar is a box constraint (helps in feature selection

Classification model selection via bilevel programming

G. KUNAPULI*†, K. P. BENNETT†, JING HU† and JONG-SHI PANG‡

Neat SVM formulation

- But how do we solve?
  - write KKT for inner problems, introduce some slack variables, get

$$
\begin{aligned}
\min\ & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{|\mathcal{N}_t|}\sum_{i\in\mathcal{N}_t}\zeta_i^t \\
\text{s. t.}\ & \lambda_{\mathbf{lb}} \le \lambda \le \lambda_{\mathbf{ub}},\ \overline{\mathbf{w}}_{\mathbf{lb}} \le \overline{\mathbf{w}} \le \overline{\mathbf{w}}_{\mathbf{ub}}, \\
& \text{and for } t = 1\ldots T, \\
& \left.\begin{array}{l} 0 \le \zeta_i^t \quad \perp y_i\left(\mathbf{x}_i'\mathbf{w}^t - b_t\right) + z_i^t \ge 0 \\ 0 \le z_i^t \quad \perp 1 - \zeta_i^t \ge 0 \end{array}\right\} \forall\, i \in \mathcal{N}_t, \\
& \left.\begin{array}{l} 0 \le \alpha_j^t \quad \perp y_j(\mathbf{x}_j'\mathbf{w}^t - b_t) - 1 + \xi_j^t \ge 0 \\ 0 \le \xi_j^t \quad \perp 1 - \alpha_j^t \ge 0 \end{array}\right\} \forall\, j \in \overline{\mathcal{N}}_t, \\
& 0 \le \gamma^{t,+} \perp \overline{\mathbf{w}} - \mathbf{w}^t \ge 0, \\
& 0 \le \gamma^{t,-} \perp \overline{\mathbf{w}} + \mathbf{w}^t \ge 0, \\
& \lambda\,\mathbf{w}^t - \sum_{j\in\overline{\mathcal{N}}_t} y_j\alpha_j^t\mathbf{x}_j + \gamma^{t,+} - \gamma^{t,-} = 0, \\
& \sum_{j\in\overline{\mathcal{N}}_t} y_j\alpha_j^t = 0,
\end{aligned}
$$

Important: this is a complementarity constraint

0 leq f $\perp$ g geq 0

means

f geq 0

g geq 0

fg=0

Seriously icky optimization problem