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).

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)

\[
\begin{align*}
\min_{x \in X} & \quad F(x, y) \\
\text{s.t.} & \quad G(x, y) \leq 0 \\
\min_{y \in Y} & \quad f(x, y) \\
\text{s.t.} & \quad g(x, y) \leq 0 \\
& \quad x, y \geq 0
\end{align*}
\]
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

\[
\begin{align*}
\text{minimize} & \quad \Theta(W, b) \\
\text{subject to} & \quad \lambda_{lb} \leq \lambda \leq \lambda_{ub}, \quad \bar{W}_{lb} \leq \bar{W} \leq \bar{W}_{ub}, \\
& \quad \text{and for } t = 1, \ldots, T, \\
& \quad (w^t, b^t) \in \arg \min_{-\bar{w} \leq w \leq \bar{w}} \left\{ \frac{\lambda}{2} \| w \|_2^2 + \sum_{j \in N_t} \max \left(1 - y_j(x_j^t w - b), 0\right) \right\}.
\end{align*}
\]

\(w^t, b^t\) are the folds, one per fold

\(W\) is a matrix of these

\(\bar{w}\) is a box constraint (helps in feature selection)
Neat SVM formulation

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

\[
\begin{align*}
\min & \quad \frac{1}{T} \sum_{t=1}^{T} \frac{1}{|\mathcal{N}_t|} \sum_{i \in \mathcal{N}_t} \zeta_i^t \\
\text{s. t.} & \quad \lambda_{lb} \leq \lambda \leq \lambda_{ub}, \quad \mathbf{w}_{lb} \leq \mathbf{w} \leq \mathbf{w}_{ub}, \\
& \quad \text{and for } t = 1 \ldots T, \\
& \quad 0 \leq \zeta_i^t \perp y_i (\mathbf{x}_i^t \mathbf{w}^t - b_t) + z_i^t \geq 0 \\
& \quad 0 \leq z_i^t \perp 1 - \zeta_i^t \geq 0 \\
& \quad 0 \leq \alpha_j^t \perp y_j (\mathbf{x}_j^t \mathbf{w}^t - b_t) - 1 + \xi_j^t \geq 0 \\
& \quad 0 \leq \xi_j^t \perp 1 - \alpha_j^t \geq 0 \\
& \quad 0 \leq \gamma^{t+} \perp \mathbf{w} - \mathbf{w}^t \geq 0, \\
& \quad 0 \leq \gamma^{t-} \perp \mathbf{w} + \mathbf{w}^t \geq 0, \\
& \quad \lambda \mathbf{w}^t - \sum_{j \in \mathcal{N}_t} y_j \alpha_j^t \mathbf{x}_j + \gamma^{t+} - \gamma^{t-} = 0, \\
& \quad \sum_{j \in \mathcal{N}_t} y_j \alpha_j^t = 0,
\end{align*}
\]

Important: this is a complementarity constraint
\[
0 \leq f \perp g \geq 0
\]

means
\[
\begin{align*}
& f \geq 0 \\
& g \geq 0 \\
& fg = 0
\end{align*}
\]

Seriously icky optimization problem