Last topic: "Interpretable duals."

Regression: want to fit a vector $b \in \mathcal{R}^m$ using as explanatory variables the columns of a matrix $A \in \mathcal{R}^{m*n}$. Want $x \in \mathcal{R}^n$ with $b - Ax$ "small".

Definition 1 *The L_p-norm of a vector $v \in \mathcal{R}^l$ is*

$$\begin{aligned} ||v||_p &= (\sum_{j=1}^{l} |v_j|^p)^{\frac{1}{p}}, \quad 1 \leq p < \infty, \\ ||v||_\infty &= \max_{1 \leq j \leq l} |v_j|. \end{aligned}$$

L_∞-regression: choose x to minimize $||Ax - b||_\infty$. Formulate this as the LP:

$$\begin{aligned} \min \quad & \beta \\ \beta e + Ax &\geq b \\ \beta e - Ax &\geq -b, \end{aligned}$$

with dual

$$\begin{aligned} \max \quad & b^T y - b^T z \\ e^T y + e^T z &= 1 \\ A^T y - A^T z &= 0 \\ y,\, z &\geq 0. \end{aligned}$$

Let $u = y - z$. Then the objective is $\max b^T u$ and the "A" constraints are $A^T u = 0$. The constraints $e^T y + e^T z = 1$ and $y \geq 0$, $z \geq 0$ imply $\sum |u_i| \leq 1$, i.e. $||u||_1 \leq 1$. Moreover, for any such u, there are suitable y and z. Hence the dual can be written

$$\begin{aligned} \max \quad & b^T u \\ A^T u &= 0 \\ ||u||_1 &\leq 1. \end{aligned}$$

L_1-regression: choose x to minimize $||Ax - b||_1$:

$$\begin{aligned} \min \quad & e^T v + e^T w \\ Ax + v - w &= b \\ v,\, w &\geq 0, \end{aligned}$$

with dual

$$\begin{aligned} \max \quad & b^T u \\ A^T u &= 0 \\ u &\le e \\ -u &\le e. \end{aligned}$$

So we get the simplified dual

$$\begin{aligned} \max \quad & b^T u \\ A^T u &= 0 \\ ||u||_\infty &\le 1. \end{aligned}$$

To treat the general L_p-case, we need conic duality: consider

$$\begin{aligned} (P) \quad \min \quad & c^T x \\ Ax &= b \\ x &\in K, \end{aligned}$$

and

$$\begin{aligned} (D) \quad \max \quad & b^T y \\ A^T y + s &= c \\ s &\in K^*. \end{aligned}$$

Here $A \in \mathcal{R}^{m*n}$, $c \in \mathcal{R}^n$, $b \in \mathcal{R}^m$. So $y \in \mathcal{R}^m$, $x, s \in \mathcal{R}^n$. K is a closed convex cone in $\mathcal{R}^n$, and $K^* = \{s \in \mathcal{R}^n : s^T x \ge 0 \text{ for all } x \in K\}$ is its dual cone. E.g., $K = \mathcal{R}^n_+$ implies $K^* = \mathcal{R}^n_+$. $K = S^{r*r}_+$ implies $K^* = S^{r*r}_+$. where S^{r*r}_+ is the set of positive semidefinite matrices of order r.

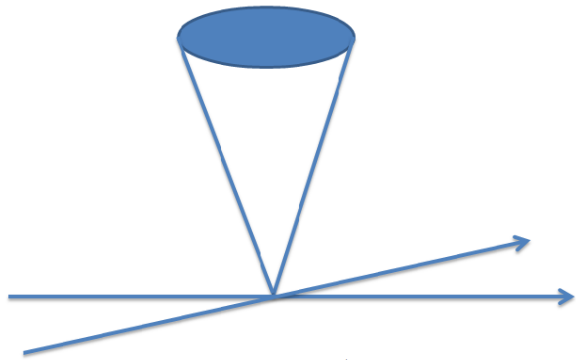
Weak duality: if x is feasible in (P), (y, s) in (D), then $c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = s^T x \ge 0$.

Definition 2 *x is a strictly feasible solution for (P) if $Ax = b$ and $x \in$ int K. Similarly, (y, s) is a strictly feasible solution for the dual if $A^T y + s = c, s \in$ int K^*.*

Theorem 1 *(Strong duality) If either (P) or (D) has a strictly feasible solution, then (P) and (D) have equal optimal values (possible infinite). If (P) ((D) resp.) has a strictly feasible solution, and (D) ((P) resp.) has a feasible solution, then (D) ((P) resp.) has a bounded nonempty set of optimal solutions.*

Proposition 1 *If K_1 and K_2 are closed convex cones in $\mathcal{R}^m$ and $\mathcal{R}^n$, then so is $K_1 \times K_2$ in $\mathcal{R}^{m+n}$, with $(K_1 \times K_2)^* = K_1^* \times K_2^*$.*

Lemma 1 *(Hölder's inequality) If $1 \le p, q \le \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, then for any $x, s \in \mathcal{R}^n$, $|s^T x| \le ||s||_p ||x||_q$. Moreover, for any x (s), there is a nonzero s (x) for which equality holds.*



$K_2^{1+2} = (K_2^{1+2})^*$

Definition 3 *Given* $1 \le p \le \infty$, *let* $K_p^{1+n} = \{(\xi, x) \in \mathcal{R}^{1+n}, \xi \ge ||x||_p\}$.

Proposition 2 *For* $1 \le p, q \le \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, $(K_p^{1+n})^* = K_q^{1+n}$.

Proof: Suppose $(\xi, x) \in K_p^{1+n}$, $(\eta, y) \in K_q^{1+n}$; then

$$\begin{aligned}\xi\eta + x^T y &\ge \xi\eta - |x^T y| \\ &\ge \xi\eta - ||x||_p ||y||_q \\ &\ge ||x||_p \eta - ||x||_p ||y||_q \\ &\ge ||x||_p (\eta - ||y||_q) \ge 0.\end{aligned}$$

Suppose $(\eta, y) \notin K_q^{1+n}$, so that $\eta < ||y||_q$. Then by the lemma, there is a nonzero x with $x^T y = -||x||_p ||y||_q$. Choose $\xi = ||x||_p$, so that $(\xi, x) \in K_p^{1+n}$. Then $\xi\eta + x^T y = ||x||_p \eta - ||x||_p ||y||_q = ||x||_p (\eta - ||y||_q) < 0$. So $(\eta, y) \notin (K_p^{1+n})^*$. $\square$

Now we can formulate L_p-regression, $\min ||Ax - b||_p$, as:

$$\begin{aligned}(P) \quad \min \quad & \beta \\ Ax + v &= b \\ (x; \beta; v) &\in \mathcal{R}^n \times K_p^{1+m},\end{aligned}$$

with dual

$$\begin{aligned}(D) \quad \max \quad & b^T u \\ A^T u + s &= 0 \\ \omega &= 1 \\ u + w &= 0 \\ (s; \omega; w) &\in \{0\} \times K_q^{1+m}.\end{aligned}$$

This gives the simplified form of the dual,

$$\begin{aligned}\max \quad & b^T u \\ A^T u &= 0 \\ ||u||_q &\le 1.\end{aligned}$$

In general, the distance b from the subspace $\{Ax\}$ in the L_p norm is the maximum component of b in a direction in the null space of A^T with L_q norm at most 1.

Slightly more complicated case: LASSO. Instead of choosing carefully a few columns of A, choose all imaginable ones, corresponding, say, to Fourier expansion, wavelets, splines,... . We want to represent b in terms of a few columns of A (avoid overfitting). As a surrogate for minimizing the number of nonzero components of x, we use the sum of the absolute values of the components. Hence we consider $\min ||Ax - b||_2 + \lambda ||x||_1$, or in conic form:

$$
\begin{array}{lrcl}
(P) & \min & \lambda\xi + \beta & \\
 & Ax + v & = & b \\
 & (\xi; x; \beta; v) & \in & K_1^{1+n} \times K_2^{1+m},
\end{array}
$$

with dual

$$
\begin{array}{lrcl}
(D) & \max & b^T u & \\
 & \sigma & = & \lambda \\
 & A^T u + s & = & 0 \\
 & \omega & = & 1 \\
 & u + w & = & 0 \\
 & (\sigma; s; \omega; w) & \in & K_\infty^{1+n} \times K_2^{1+m},
\end{array}
$$

or in simpler terms,

$$
\begin{array}{rcl}
\max & b^T u & \\
||u||_2 & \leq & 1 \\
||A^T u||_\infty & \leq & \lambda.
\end{array}
$$