Last topic: “Interpretable duals.”

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

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

$$||v||_p = \left( \sum_{j=1}^{l} |v_j|^p \right)^{\frac{1}{p}}, \quad 1 \leq p < \infty,$$

$$||v||_{\infty} = \max_{1 \leq j \leq l} |v_j|.$$

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

$$\min \quad \beta$$

$$\beta e + Ax \geq b$$

$$\beta e - Ax \geq -b,$$

with dual

$$\max \quad b^T y - b^T z$$

$$e^T y + e^T z = 1$$

$$A^T y - A^T z = 0$$

$$y, z \geq 0.$$

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

$$\max \quad b^T u$$

$$A^T u = 0$$

$$||u||_1 \leq 1.$$

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

$$\min \quad e^T v + e^T w$$

$$Ax + v - w = b$$

$$v, w \geq 0.$$
with dual
\[
\begin{align*}
\max & \quad b^Tu \\
A^Tu &= 0 \\
& \quad u \leq e \\
& \quad -u \leq e.
\end{align*}
\]
So we get the simplified dual
\[
\begin{align*}
\max & \quad b^Tu \\
A^Tu &= 0 \\
||u||_\infty &\leq 1.
\end{align*}
\]
To treat the general $L_p$-case, we need conic duality: consider
\[
(P) \quad \min \quad c^Tx \\
Ax &= b \\
x &\in K,
\]
and
\[
(D) \quad \max \quad b^Ty \\
A^Ty + s &= c \\
s &\in K^*.
\]
Here $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. So $y \in \mathbb{R}^m$, $x, s \in \mathbb{R}^n$. $K$ is a closed convex cone in $\mathbb{R}^n$, and $K^* = \{s \in \mathbb{R}^n : s^T x \geq 0 \text{ for all } x \in K\}$ is its dual cone. E.g., $K = \mathbb{R}_+^n$ implies $K^* = \mathbb{R}_+^n$. $K = S^{r \times r}_+$ implies $K^* = S^{r \times r}_+$. where $S^{r \times r}_+$ is the set of positive semidefinite matrices of order $r$.

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

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

**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 $\mathbb{R}^m$ and $\mathbb{R}^n$, then so is $K_1 \times K_2$ in $\mathbb{R}^{m+n}$, with $(K_1 \times K_2)^* = K_1^* \times K_2^*$.

**Lemma 1** (Hölder’s inequality) If $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, then for any $x, s \in \mathbb{R}^n$, $|s^T x| \leq ||s||_p ||x||_q$. Moreover, for any $x$ (s), there is a nonzero $s$ (x) for which equality holds.
Definition 3 Given $1 \leq p \leq \infty$, let $K_p^{1+n} = \{(\xi, x) \in \mathcal{R}^{1+n}, \xi \geq ||x||_p\}$.

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

Proof: Suppose $(\xi, x) \in K_p^{1+n}, (\eta, y) \in K_q^{1+n}$; then
\[
\begin{align*}
\xi \eta + x^T y & \geq \xi \eta - |x^T y| \\
& \geq \xi \eta - ||x||_p ||y||_q \\
& \geq ||x||_p \eta - ||x||_p ||y||_q \\
& \geq ||x||_p (\eta - ||y||_q) \geq 0.
\end{align*}
\]
Suppose $(\eta, y) \not\in K_q^{1+n}$, 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 $(\eta, y) \not\in (K_p^{1+n})^*$. \(\square\)

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

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

with dual

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

This gives the simplified form of the dual,
\[
\begin{align*}
\max \quad & b^T u \\
A^T u & = 0 \\
||u||_q \leq 1.
\end{align*}
\]
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{align*}
(P) \quad & \min \lambda \xi + \beta \\
& Ax + v = b \\
& (\xi; x; \beta; v) \in K_1^{1+n} \times K_2^{1+m},
\end{align*}

with dual

\begin{align*}
(D) \quad & \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{align*}

or in simpler terms,

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