Lasso Screening

October 9, 2024

1 Screening

Let f be differentiable and strictly convex, let $X \in \mathbb{R}^{n \times p}$, $\lambda > 0$. Consider

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

The KKT conditions are

$$\begin{cases} X_j^T(y - X\beta^*) = \lambda \cdot \text{sgn}(\beta_j^*) & \beta_j^* \neq 0 \\ |X_j^T(y - X\beta^*)| \leq \lambda & \beta_j^* = 0 \end{cases}$$

which implies that

$$|X_j^T(y - X\beta^*)| < \lambda \quad \Longrightarrow \quad \beta_j^* = 0 \tag{1}$$

Suppose that the dual solution is $u^\star$. Then $\beta^\star, z^\star$ must minimize $L(\beta, z, u^\star)$

$$
\begin{aligned}
&\nabla_z L(\beta, z, u^\star) = 0 \\
\Longleftrightarrow &\nabla_z \left\{ \frac{1}{2} \|y - z\|_2^2 + u^\top z \right\} = 0 \\
\Longleftrightarrow & -(y - z^\star) + u^\star = 0 \\
\Longleftrightarrow &y - X\beta^\star = u^\star
\end{aligned}
$$

Replace $y - X\beta^\star$ using u^* in (1) we get

$$
|X_j^T u^*| < \lambda \quad \Longrightarrow \quad \beta_j^* = 0.
$$

This is when u^* is in the interior of the slab defined by the feature X_j.

2 SAFE Rule

Assume all data are normalized $\|y\| = 1$ and $\|X_j\| = 1$ for $j = 1, \ldots, p$. The lasso dual is equivalent to

$$
\begin{aligned}
&\min_u \|\frac{y}{\lambda} - u\|_2^2 \\
&\text{subject to} \|X^\top u\|_\infty \leq 1
\end{aligned} \tag{2}
$$

where

$$
X\beta^\star = y - \lambda u^\star
$$

and

$$
X_j^\top u^* = \begin{cases} \operatorname{sgn}\left(\beta_j^*\right) & \text{if } \beta_j^* \neq 0 \\ [-1, 1] & \text{if } \beta_j^* = 0 \end{cases} . \tag{3}
$$

Since $\|y\|_2^2 = 1$ and $\|X_j\|_2^2 = 1$, y and X_j lie on the unit sphere S^{p-1}. For y on S^{p-1}, the function of the sphere is $f(y) = \|y\|_2^2 = 1$, the normal to the surface is

$$\nabla_y f(y) = \nabla_y \|y\|_2^2 = 2y$$

thus at point y of the sphere S^{p-1}, the tangent hyperplane is

$$\nabla_y f(y)^\top (z - y) = 0 \Longleftrightarrow y^\top (z - y) = 0 \Longleftrightarrow y^\top z = 1$$

Thus we denote this tangent hyperplane at y by

$$P(y) = \{z : z^\top y = 1\}.$$

And we denote

$$H(y) = \{z : z^\top y \leq 1\}$$

the corresponding closed half space containing the origin. The constraints in (2) indicate that feasible u must be in $H(X_j)$ and $H(-X_j)$ for all j. To find u^* that minimizes the objective in (2), we must find a feasible u that is closest to y/λ.

If u^* is not on $P(X_j)$ or $P(-X_j)$, then $\beta_j^* = 0$ and we can safely discard X_j from the problem. Now let $\lambda_{\max} = \max_j |X_j^\top y|$ and $X_{j'}$ be selected so that $\lambda_{\max} = |X_{j'}^\top y|$. Note that $u' = y/\lambda_{\max}$ is a feasible solution for (2), so that u' must be in the polyhedron, since

$$|X_j^\top u'| = \frac{|X_j^\top y|}{\lambda_{\max}} = \frac{|X_j^\top y|}{\max_j |X_j^\top y|} \leq 1 \qquad \forall j$$

$\lambda_{\max}$ is also the smallest λ for which the primal problem has zero solution. If $\lambda > \lambda_{\max}$, then y/λ itself is feasible, making it the optimal solution. Since it is not on any hyperplane $P(X_j)$ or $P(-X_j)$, $\beta_j^* = 0$ for $j = 1, \ldots, p$. Hence we assume that $\lambda \leq \lambda_{\max}$.

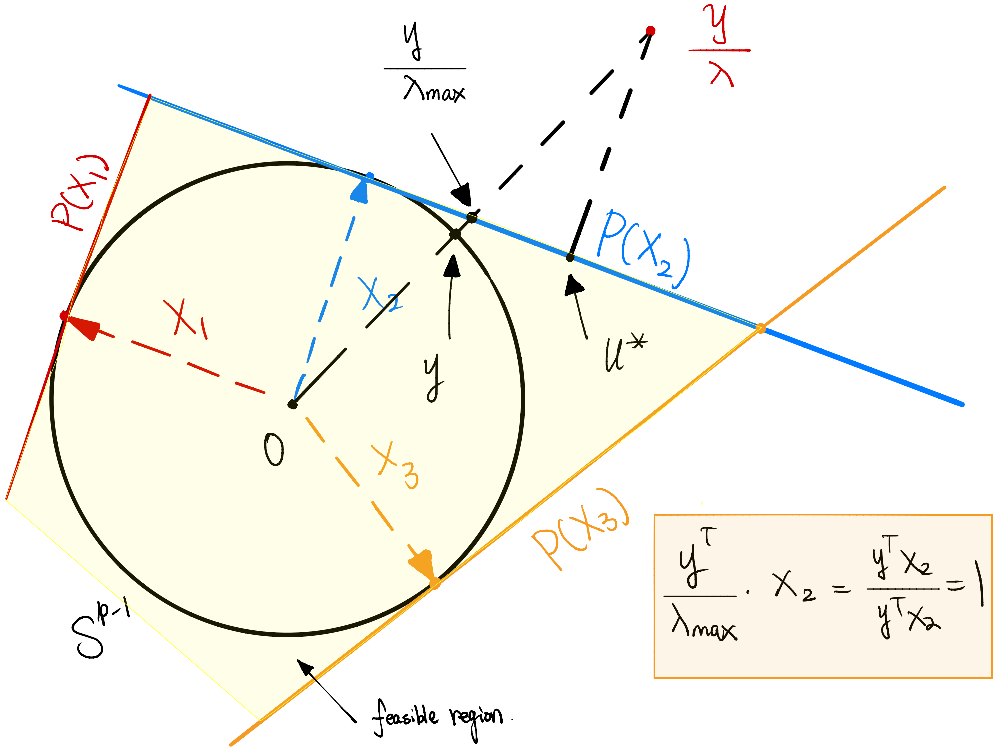

Figure 1: Geometry of the dual problem.

These observations can be used for screening as follows. If we know that u is within a region $\mathcal{R}$, then we can discard those X_j for which the tangent hyperplanes $P(X_j)$ or $P(-X_j)$ don't intersect $\mathcal{R}$, since by (3) the corresponding $\beta_j = 0$. Moreover, if the region $\mathcal{R}$ is contained in a closed ball centered at q with radius r

$$\{u : \|u - q\|_2 \leq r\}$$

then one can discard all X_j for which $|X_j^\top q|$ is smaller than a threshold determined by the common tangent hyperplanes of the spheres $\|u - q\|_2 = r$ and S^{p-1}. This "sphere test" is made precise in the following lemma

Lemma 1. *If the solution* u^* *of (2) satisfies* $\|u^* - q\|_2 \leq r$, *then* $|X_j^\top q| < (1 - r) \implies \beta_j^* = 0$.

Proof of Lemma 1. Assume that we have $|X_j^\top q| < (1 - r)$. According to (3), in order to assert that $\beta_j = 0$, we only need to prove that for the optimal solution u^* of (3): $|X_j^\top u^*| < 1$, which can be proved by:

$$\begin{aligned} |X_j^\top u^*| &= |X_j^\top (u^* - q) + X_j^\top q| \\ &\leq |X_j^\top (u^* - q)| + |X_j^\top q| \\ &\leq \|X_j\|_2 \|u^* - q\|_2 + |X_j^\top q| \\ &< r + (1 - r) \\ &= 1 \end{aligned}$$

□

El Ghaoui's SAFE rule is a sphere test of the simplest form. To see this, note that $y/\lambda_{\max}$ is a feasible point of (2), so the optimal u^* cannot be further away from y/λ than $y/\lambda_{\max}$ from y/λ. Therefore we have the constraint:

$$\|u - \frac{y}{\lambda}\|_2 \leq \|\frac{y}{\lambda} - \frac{y}{\lambda_{\max}}\|_2 \leq \left(\frac{1}{\lambda} - \frac{1}{\lambda_{\max}}\right) \|y\|_2 \leq \frac{1}{\lambda} - \frac{1}{\lambda_{\max}}.$$

Plugging in $q = y/\lambda$ and $r = 1/\lambda - 1/\lambda_{\max}$ into Lemma 1 yields El Ghaoui's SAFE rule.

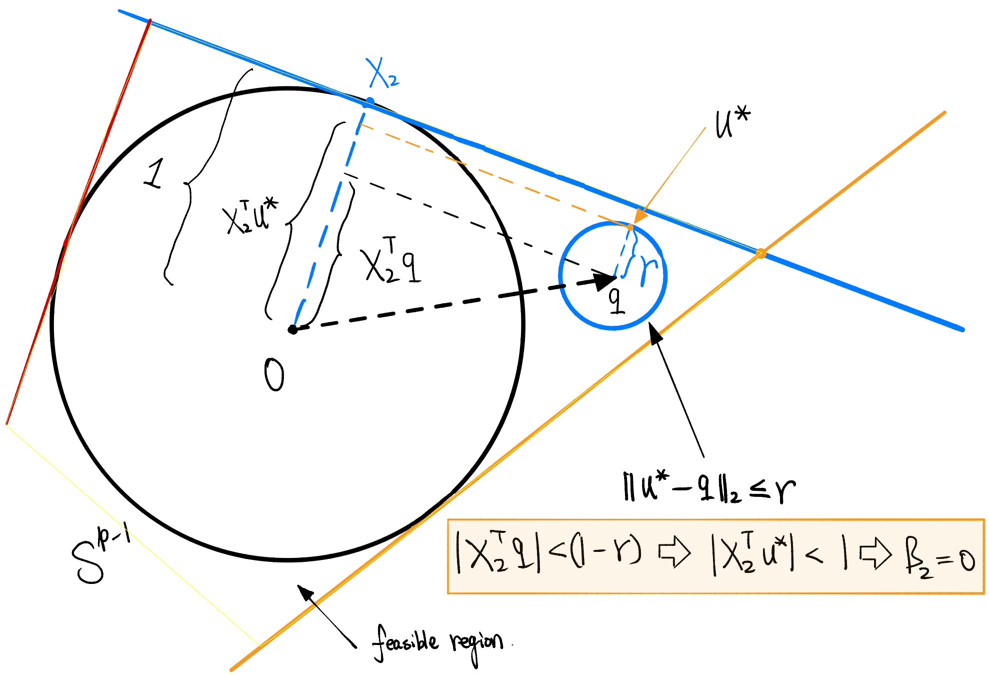

Figure 2: Illustration of a sphere test.

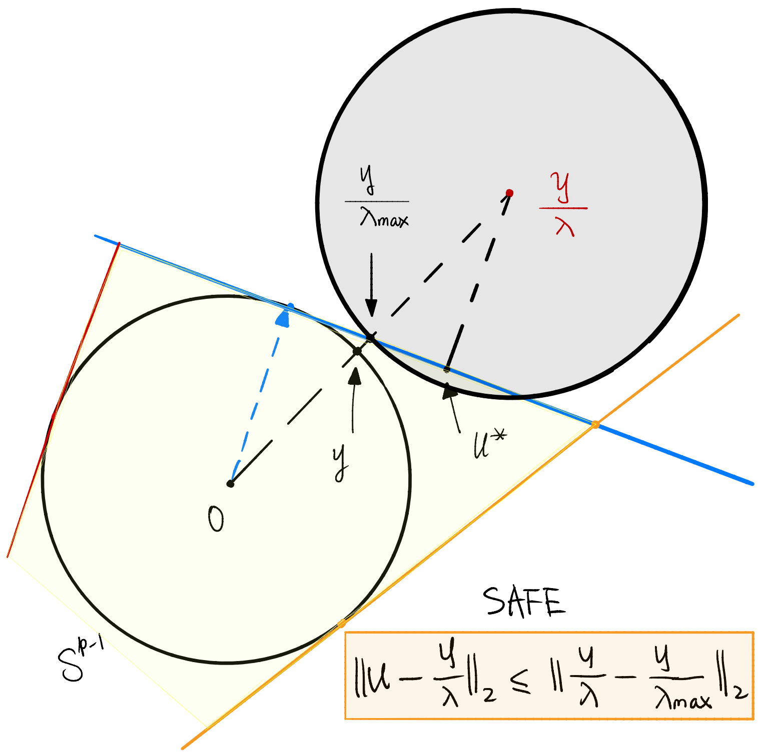

Figure 3: The solid red circles leading to sphere tests SAFE