为什么 ridge 不能把 β 减小到 0，而 Lasso 可以？

例子： $y = x\beta + \varepsilon$，一元模型.

$$\text{Lasso}: \min_{\beta} \sum_{i=1}^{n} (y - x\beta)^2 + \lambda|\beta| = y'y - 2x'y\beta + \beta x'x\beta + \lambda|\beta|$$

假设 OLS 的解 $\hat{\beta} > 0$（也即 $x'y > 0$ 因为 $\hat{\beta} = (x'x)^{-1}x'y$）

因为是一元的.

① 若 Lasso 的解 $\tilde{\beta} > 0$，

$$\frac{\partial L}{\partial \beta} = -2x'y + 2\beta x'x + \lambda \Rightarrow \tilde{\beta} = \frac{2x'y - \lambda}{2x'x}$$

$$= \frac{x'y - \frac{\lambda}{2}}{x'x} \quad ①$$

增大 λ，则当 $\lambda = 2x'y$ 时，$\tilde{\beta} = 0$.

但是 λ 继续增大不能让 $\tilde{\beta}$ 变负.

因为假若 $\tilde{\beta}$ 为负，则

$$\text{Loss} = y'y - 2x'y\beta + \beta x'x\beta - \lambda\beta$$

此时 $\tilde{\beta} < 0$ 同时让 $\lambda|\beta|$ 大，也同时让 MSE 大，很显然不是最优解.

$$\frac{\partial L}{\partial \beta} = -2x'y + 2\beta x'x - \lambda = 0$$

$$\Rightarrow \tilde{\beta} = \frac{x'y + \frac{\lambda}{2}}{x'x} \quad ②$$

这时 β 的选择和 λ 增大矛盾.

由①和②可以看得清楚，$\tilde{\beta}$ 的符号会和 $\hat{\beta}$ 一样. 若 $x'y > 0$，则 $\tilde{\beta} \geq 0$；$x'y < 0$，$\tilde{\beta} \leq 0$. $\tilde{\beta}$ 不会因为 λ 变大而变符号，最多就是到 0，并且停在那里.

如果是 Ridge regression:

Loss: $y'y - 2x'y\beta + \beta x'x\beta + \lambda\beta^2$

$$\frac{\partial L}{\partial \beta} = -2x'y + 2x'x\beta + 2\lambda\beta$$

$$\beta = \frac{x'y}{x'x+\lambda}$$

因此 Ridge regression 不能得到 $\beta=0$ 的结果

如果是多维，则可能出现有些 β 变了之后导致另外的 β 变符号.