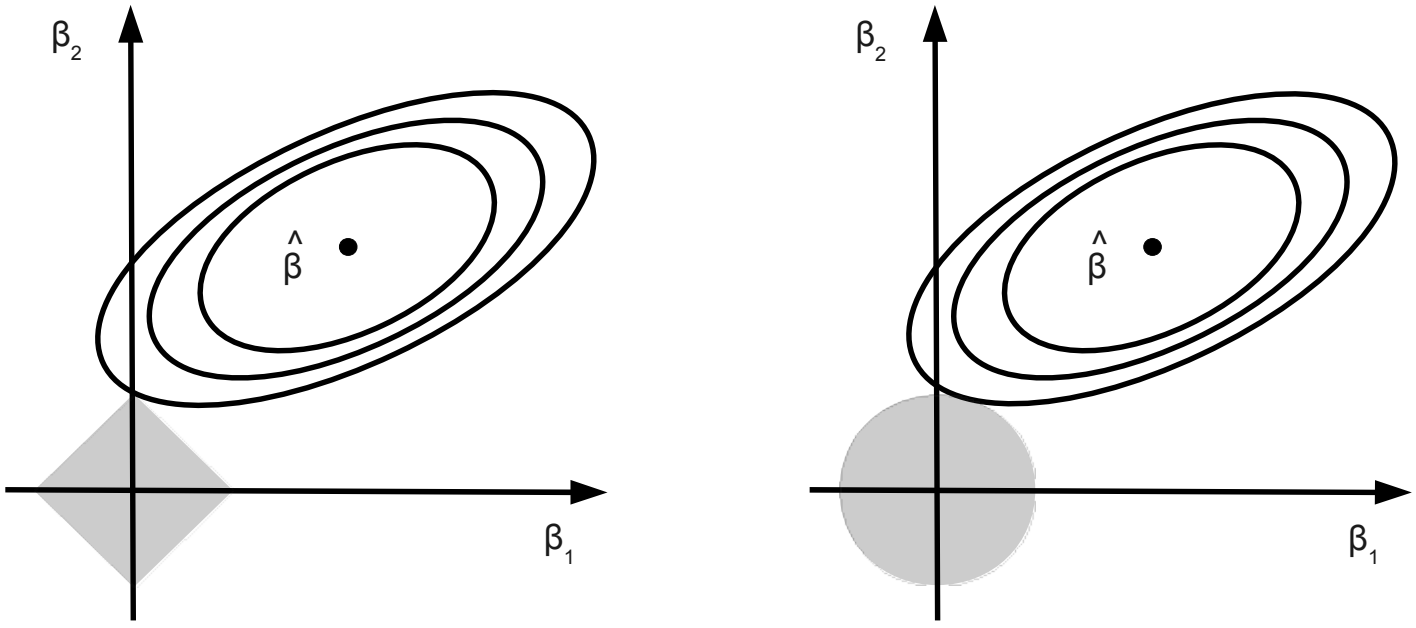
Lasso is sparse estimator

Lasso and Ridge in primal problem optimization form:



$\rightsquigarrow \hat{\beta}_j(\lambda) = 0$ for $j = 1$ and a certain λ

and this is **not** the case for Ridge regression

Lasso involves convex optimization

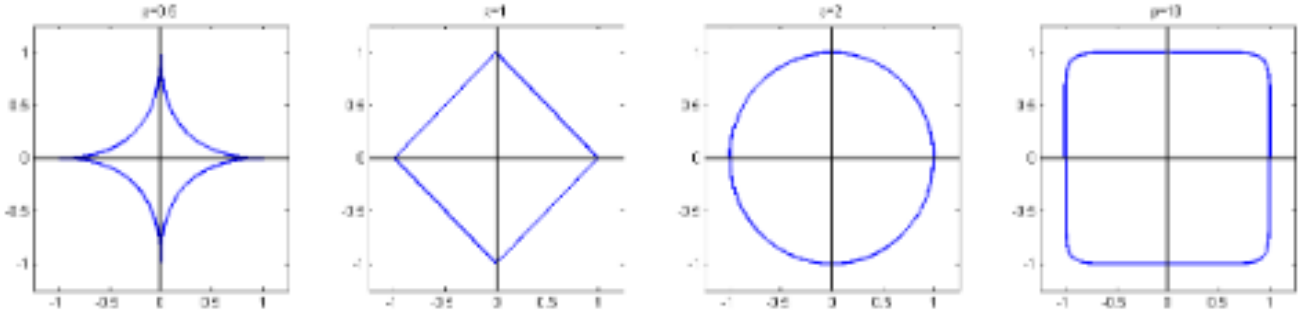


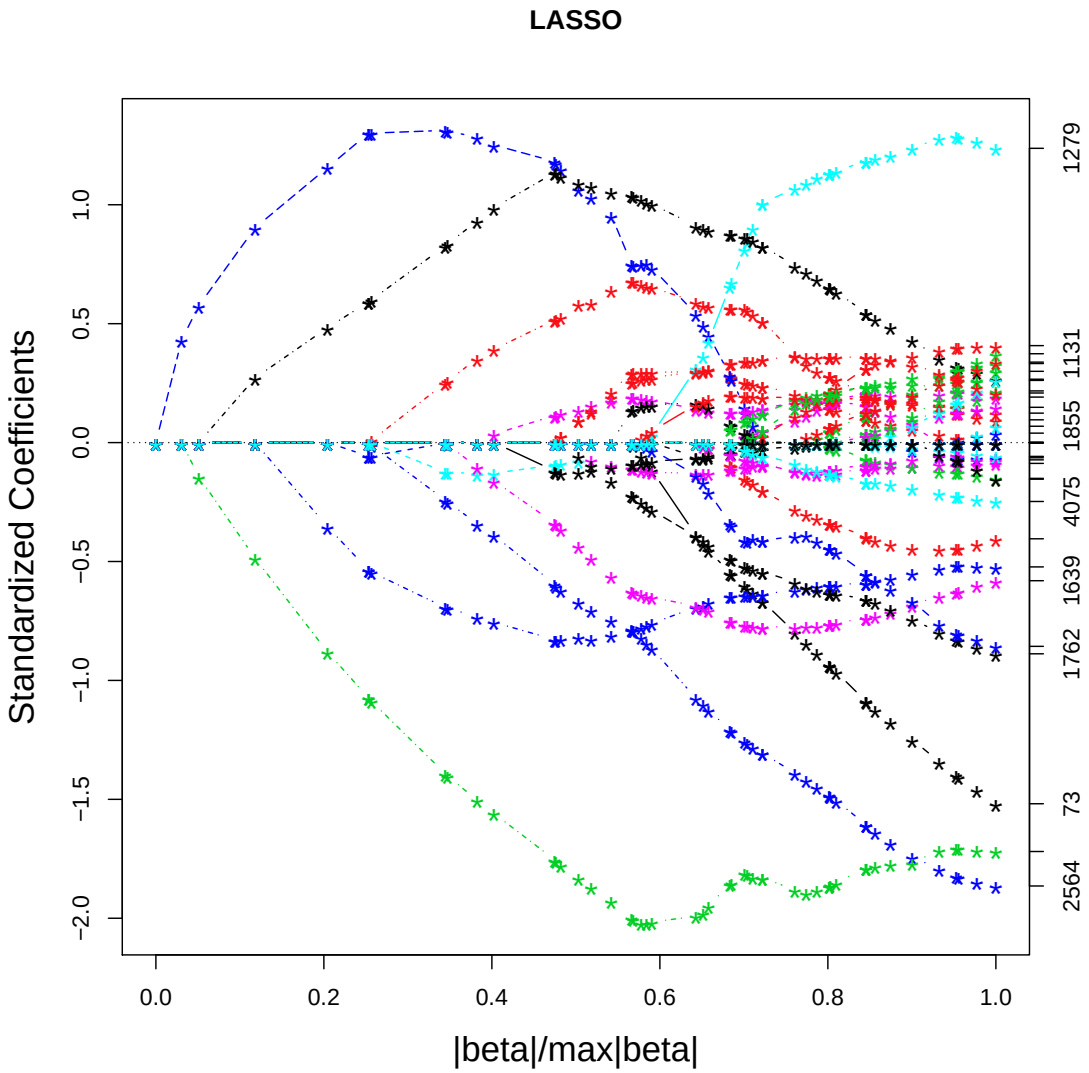
Figure 1: Unit circles for several Minkowski-p-norms $\|\mathbf{x}\|_p$: from left to right $p = 0.5$, $p = 1$ (Manhatten), $p = 2$ (Euclidean), $p = 10$.

from Lange, Zühkle, Holz, Villmann (2014)

- convex: ℓ_p-norm with $p \geq 1$
- sparse: ℓ_p-norm with $p \leq 1$ (need "edges in the ball)

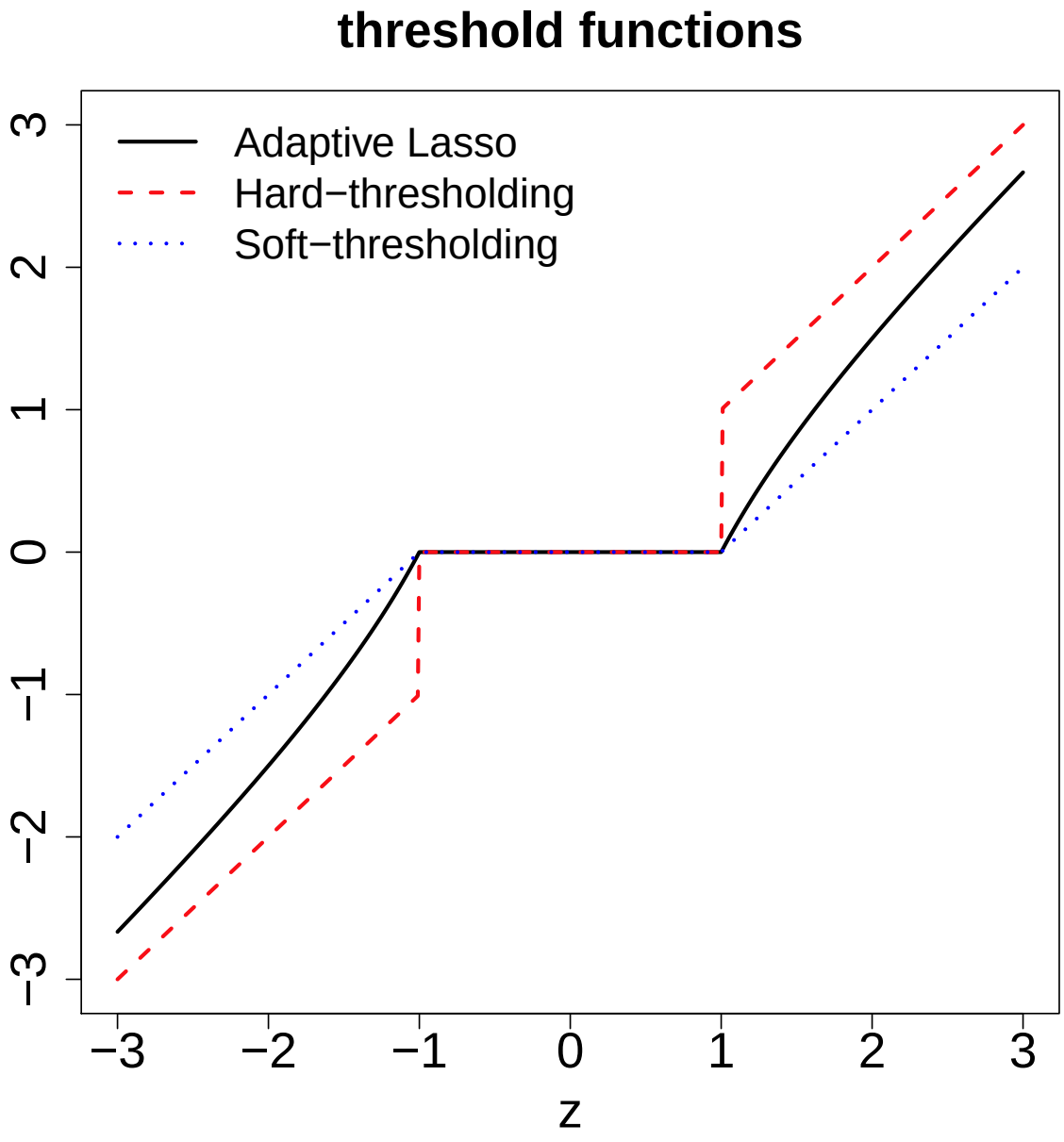
$\rightsquigarrow$ $p = 1$, that is Lasso, for sparse and convex estimator

Lasso regularization path for Riboflavin data



certainly sparse, even when $\lambda \to 0$ (x-axis is like inverse of λ)

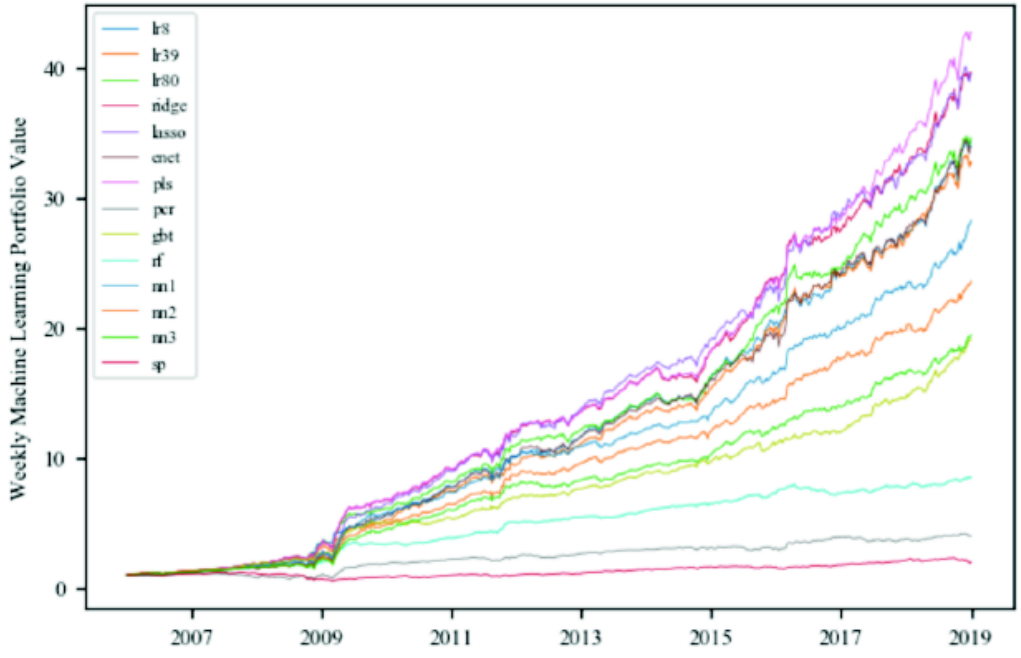
Orthonormal design: threshold functions



Lasso as a prediction machine/method

Fig. 6. Cumulative Performance for Weekly Machine Learning Portfolios

This figure depicts the cumulative weekly returns for equal-weighted long-short machine learning portfolios for 13 machine learning models. 'lr80' denotes the linear regression model with all 80 variables. 'lr39' denotes the linear regression model with variables that have strong predictability for weekly returns. 'lr8' denotes the linear regression model with the eight strongest predictor variables. 'ridge' denotes the ridge regression model. 'lasso' denotes the lasso regression model. 'enet' denotes the elastic net regression model. 'pls' denotes the partial least squares regression model. 'pcr' denotes the principal component regression model. 'gbt' denotes the gradient boosting regression tree model. 'rf' denotes the random forest regression model. 'nn1' denotes the neural network model with one hidden layer. 'sp' denotes the benchmark S&P 500 index. The accumulation period is from January 2006 to December 2018, and the initial investment is set as 1 at the start of the accumulation period.



taken from MSc thesis Jiawen Le (September 2019)