Incorporating known risk factors into models

Jiaxuan Wang

<2017-02-07 Tue>

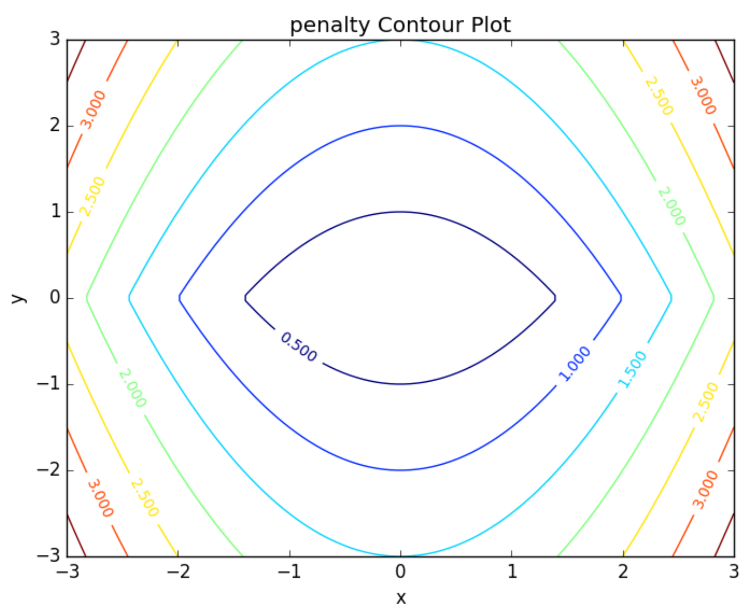
1 objective

Incorporating known risk factors with unknown risk factors in predicting outcome. In the case of choosing between correlated variables, the model should favor known risk factors.

2 approaches taken

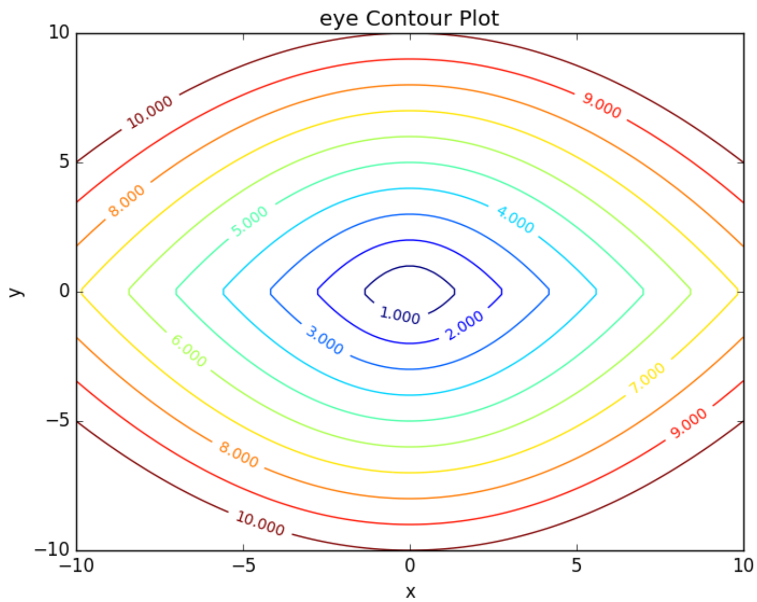
2.1 old approach

$0.5 * \lambda_2 \; ||\mathrm{r} * \theta||_2^2 + \lambda_1 \; ||(1\text{-}\mathrm{r}) * \theta||_1$ where $\mathrm{r} \in \{0,1\}^{\mathrm{d}}$, $\theta \in \mathrm{R}^{\mathrm{d}}$



2.2 new approach

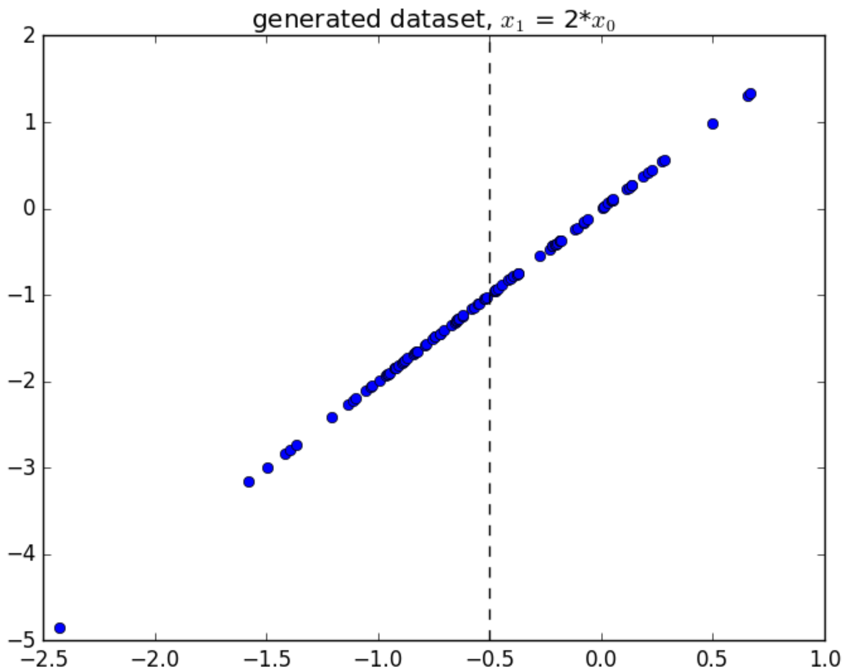
fix a convex body that have property of the previous contour plot such that the angle at the end point is 45 degree. The following is the contour plot of its induced norm



3 experiments

3.1 set up

Data n=100:



x_0 ~ N(-0.5, 0.5)

$x_1 = 2 * x_0$

Loss function is the negative loss likelihood of the logistic regression model.

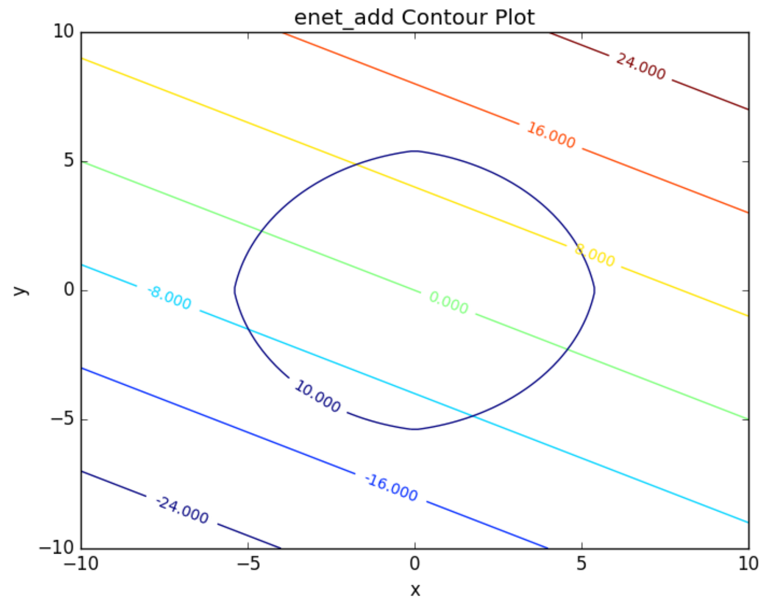
Optimizer: AdaDelta

Number of Epoch: 1000

Regulizers: elastic net, lasso, ridge, OWL, weighted lasso, weighted ridge, our penalty

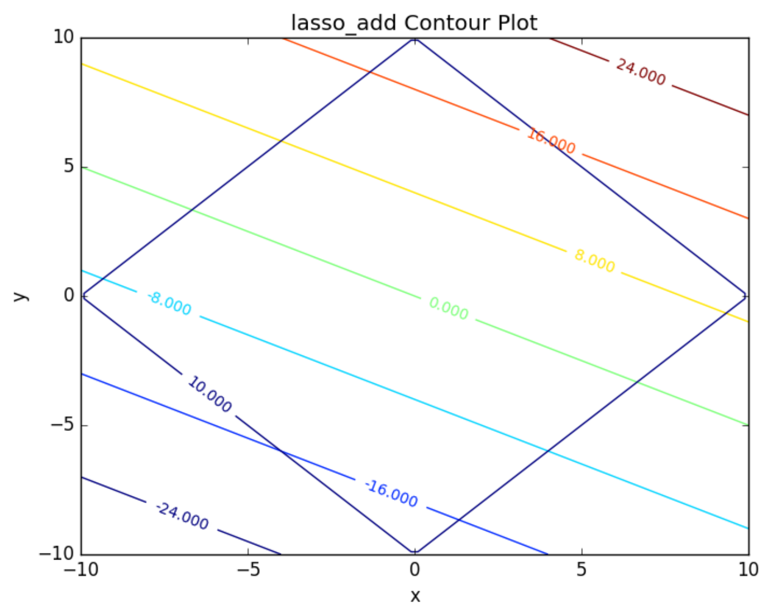
3.1.1 elastic net

α * (c * $||\theta||_1$ + 0.5 * (1 - c) * $||\theta||_2^2$) where c is a scaler



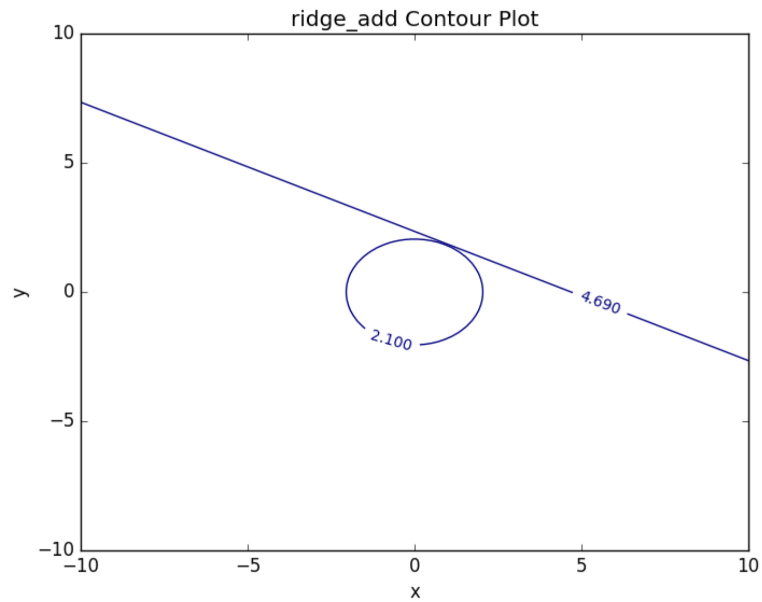
3.1.2 lasso

$\alpha * ||\theta||_1$



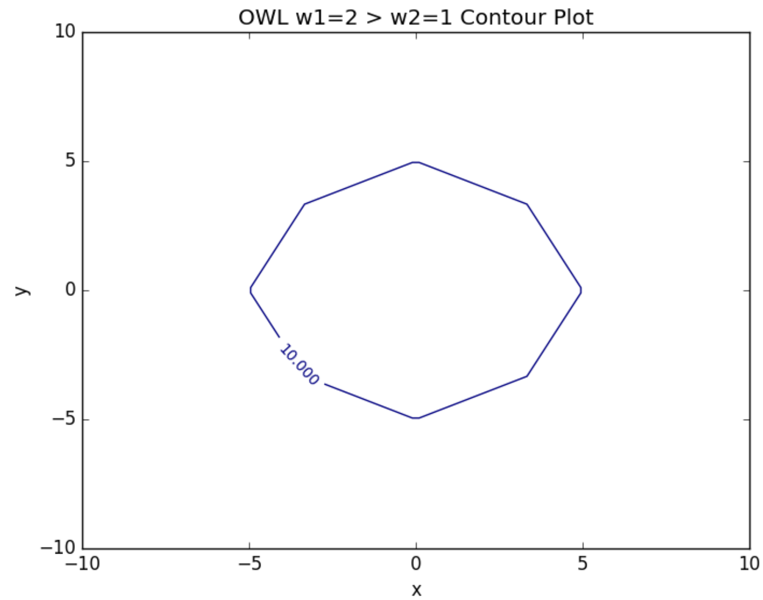
3.1.3 ridge

$0.5 * \alpha * ||\theta||_2^2$

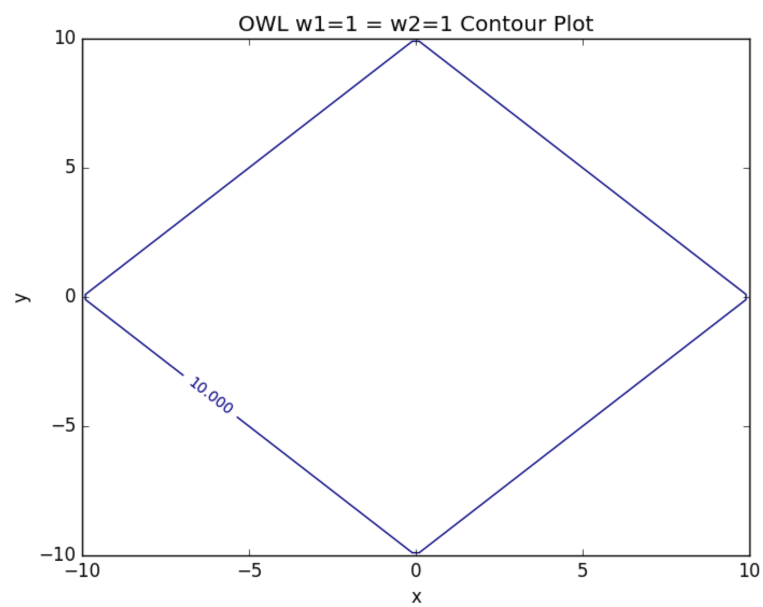


3.1.4 OWL

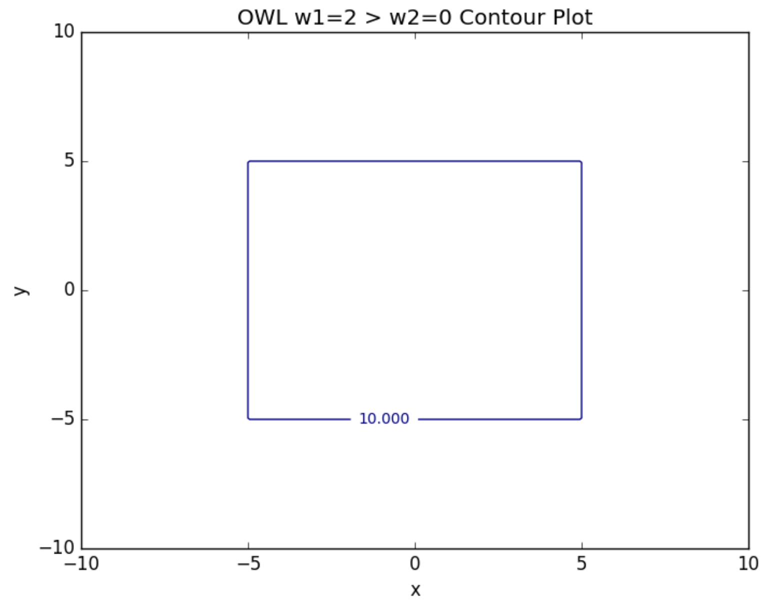
$\alpha * \sum_{i=1}^{n} w_i\ |x|_{[i]}$ where $w \in K_{m+}$ (monotone nonnegative cone)



w1=2 > w2=1.png
degenerated case: back to lasso



w1=1 = w2=1.png
degenerated case: back to l_{inf}



w1=2 > w2=0.png

some properties:

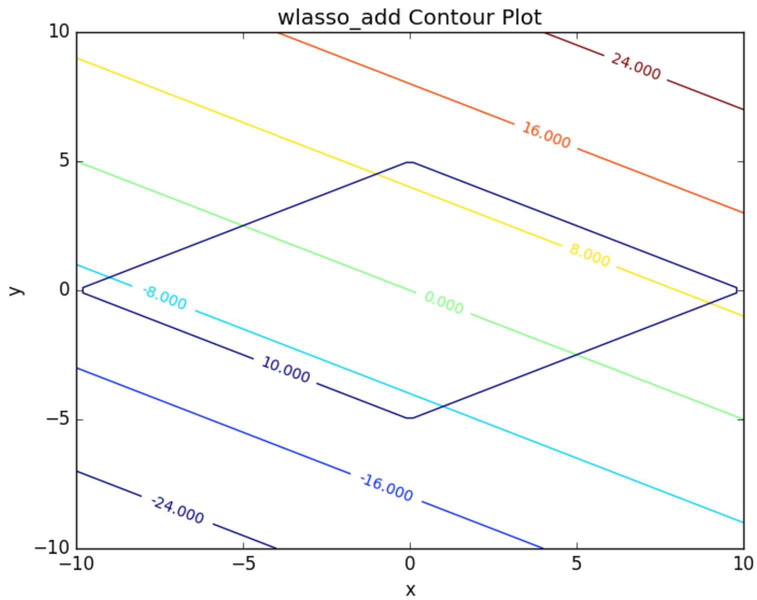
generalization of OSCAR norm

symmetry with respect to signed permutations

in the regular case, the minimal atomic set for this norm is known (the corners are easily calculated)

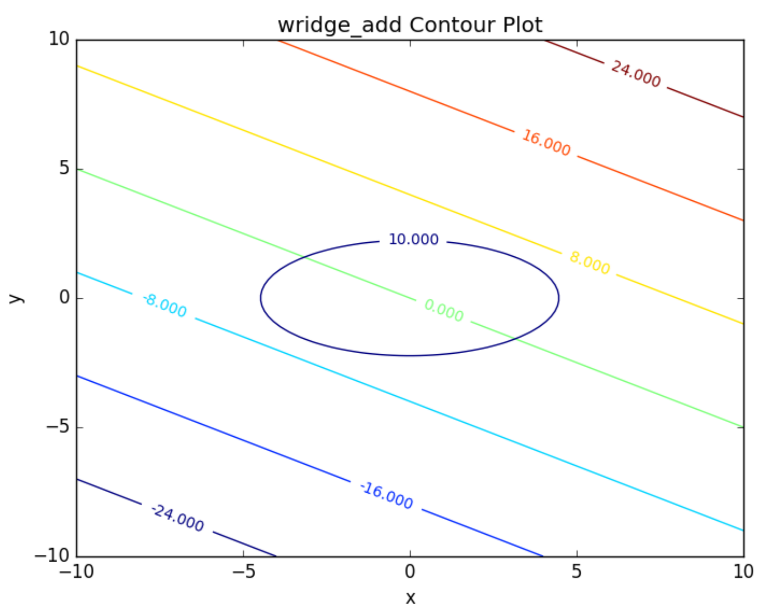
3.1.5 weighted lasso

α * $||$w * $\theta||_1$ where w $\in$ R_+$^{\text{d}}$



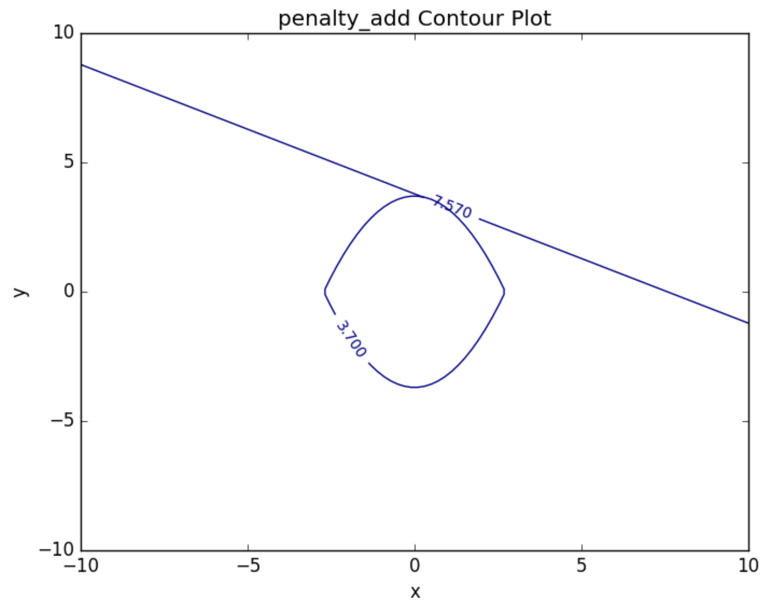
3.1.6 weighted ridge

$0.5 * \alpha * ||\mathrm{w} * \theta||_2^2$ where $\mathrm{w} \in \mathrm{R}_+^{\mathrm{d}}$



3.1.7 our penalty

α * (0.5 * (1-c) * $||r * \theta||_2^2$ + c * $||(1-r) * \theta||_1$) where r $\in \{0,1\}^d$, $\theta \in R^d$, $\alpha \in R$, c $\in$ R



3.2 running procedure

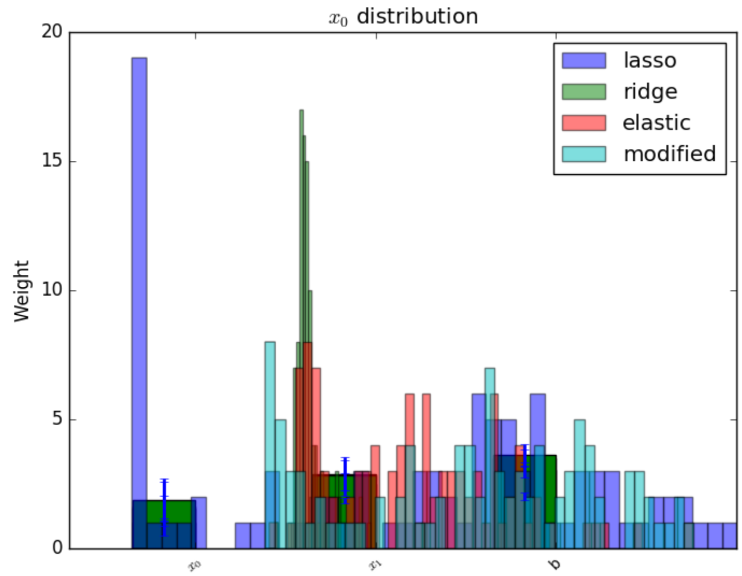
3.2.1 first run

b regularized

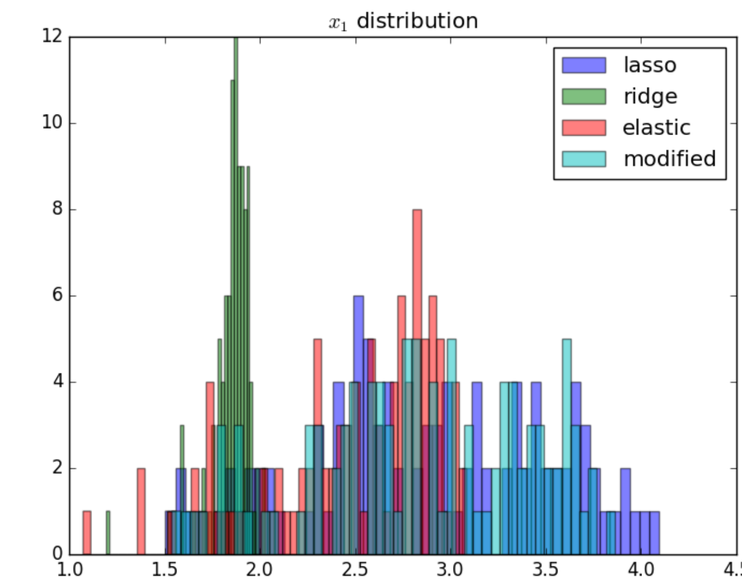
fix hyperparmeters to predefined value

repeat the following 100 times:

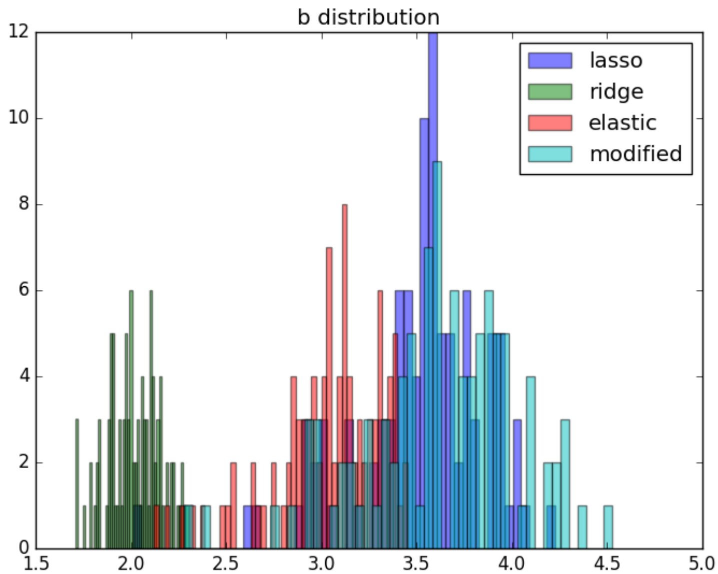
generate data, run the selected regularizers, record θ



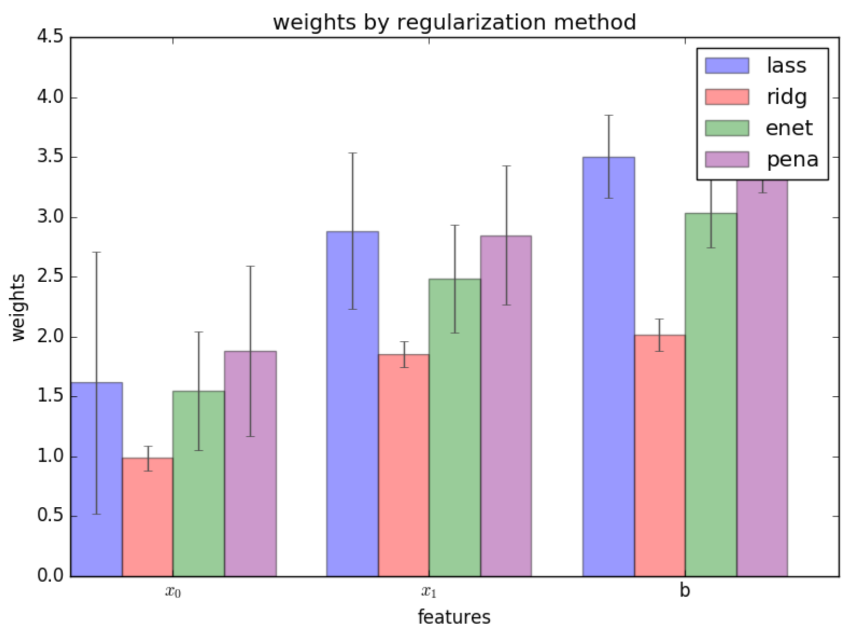
distribution.png



distribution.png



distribution.png



3.2.2 second run

b unregularized

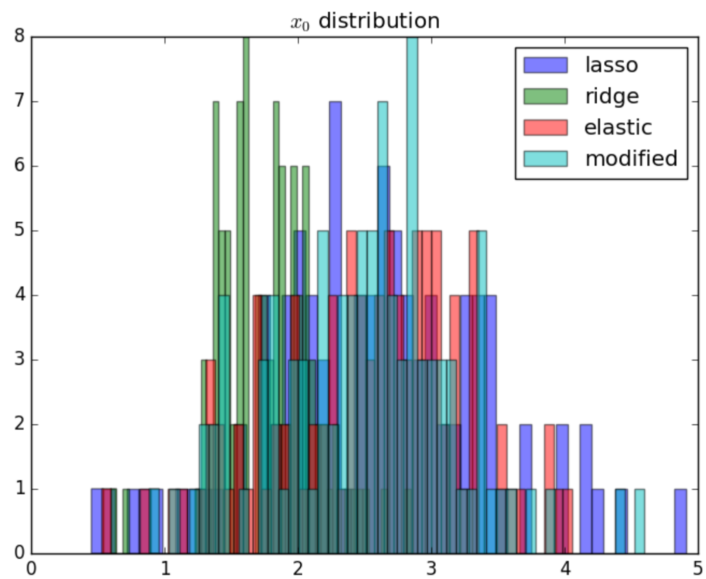
generate two datasets, one for training, one for validation

parameter search over the different hyperparams of the regularizers

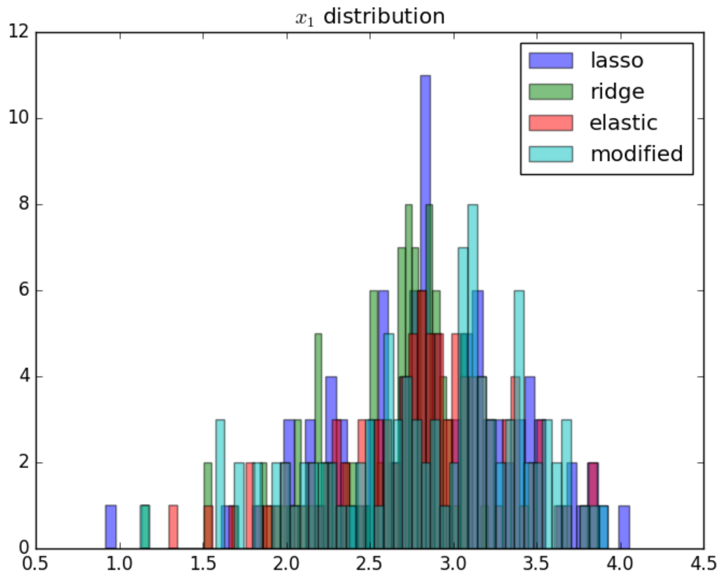
for each regularizer, use the hyperparmeters that acheives the minimal loss

repeat the following 100 times:

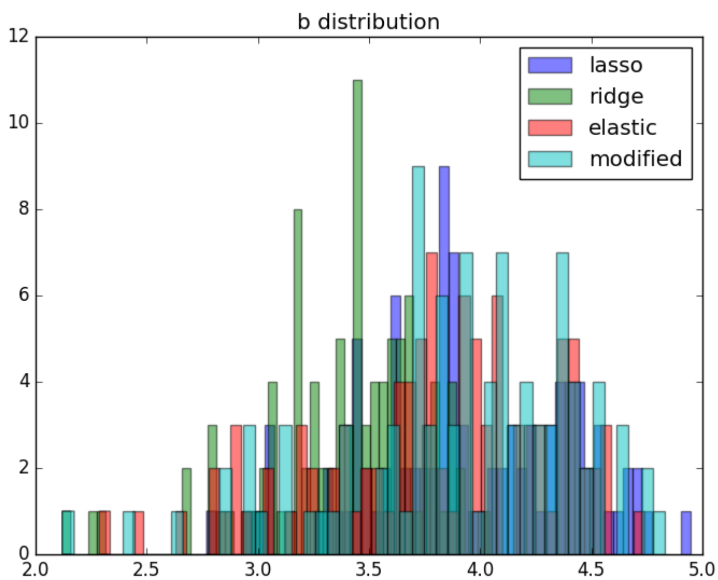
generate data, run the selected regularizers, record θ



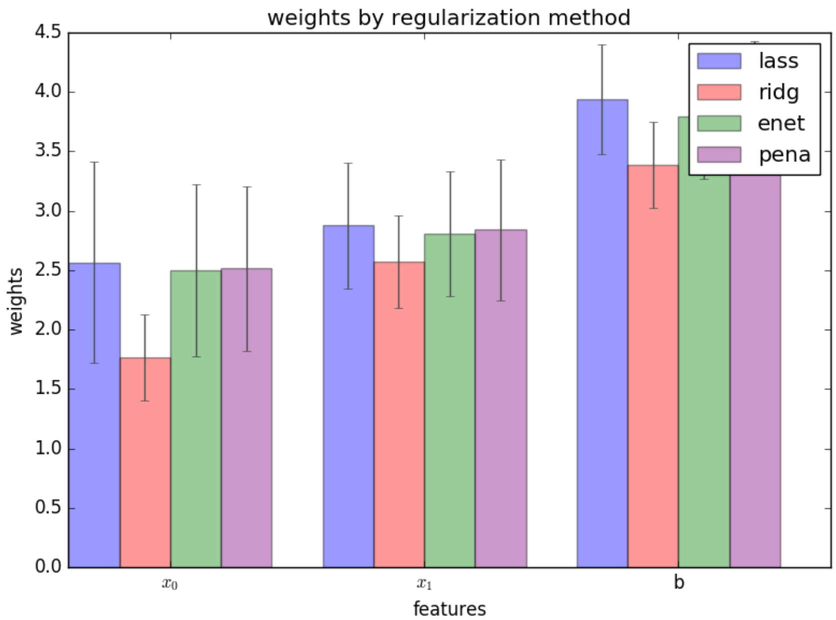
distribution.png



distribution.png



distribution.png



3.2.3 third run

b unregularized

generate two datasets, one for training, one for validation

normalize the data to zero mean and unit variance (validaton data is normalized using mean and variance for the training data)

parameter search over the different hyperparams of the regularizers

for each regularizer, use the hyperparmeters that acheives the minimal loss

repeat the following 100 times:

generate data, normalize data, run the selected regularizers, record θ