Lasso Regression

Thursday, June 10, 2021 6:42 AM

$y = mx + b$ $y = b$

$\downarrow$ $\lambda \uparrow$ under

L1 Regularization | overfitting → L2 reg

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \|w\|^2$$

$\boxed{\lambda \geq 0}$

↳ $\lambda (w_1^2 + w_2^2 + \dots + w_n^2)$

→ 0 $w_1 \to w_n \to$ coeff

overfitting | under

alpha

↳ Lasso L1 norm $\lambda [|w_1| + |w_2| + |w_3| + \dots + |w_n|]$

$$L = MSE + \boxed{\lambda \|w\|}$$ → code implement (sklearn)

→ Key point →

→ Difference Ridge Vs Lasso

underfitting

$\lambda = 0$ → Linear

overfitting

degree

x_1 x_2 x_3 0 0 x_n y

0.5 0.2

↳ ridge

↳ lasso → λ → ↑

dim <

feature selection

0

decrease

lasso

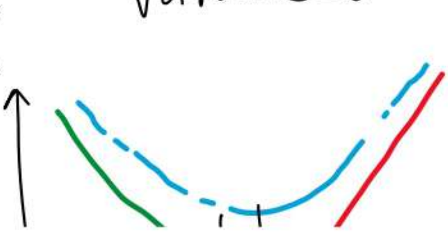
$\lambda \uparrow$

↳ overfitting ↓ Bias ↑

Variance ↓

Varince ↑ Bias ↓

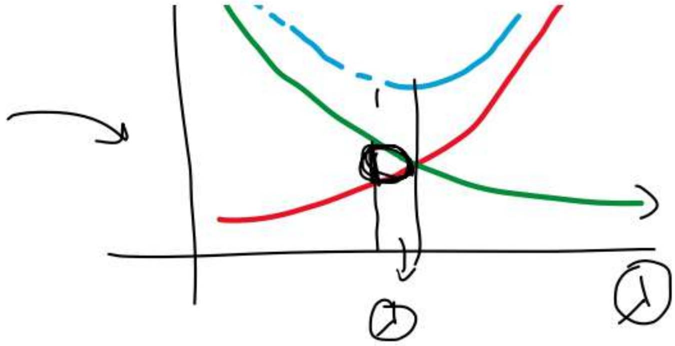
values



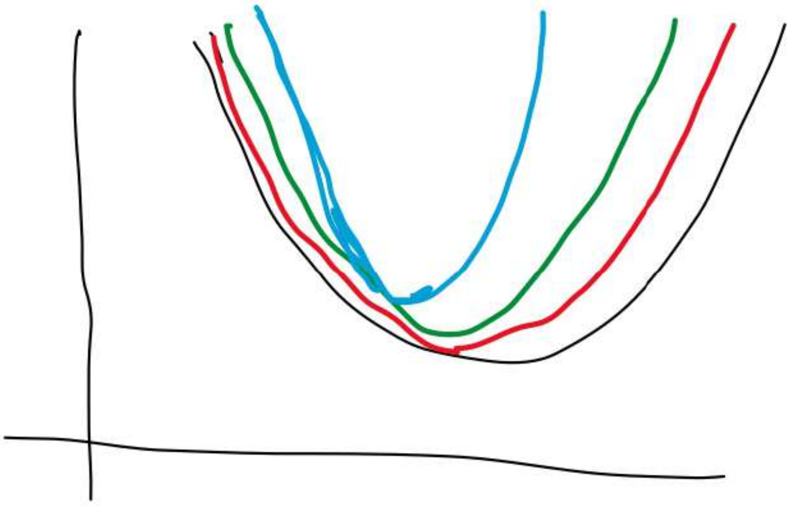
Red → bias

green - variance

blue - total error



y:
blue – total error



Understanding Sparsity

Friday, June 11, 2021 11:21 AM

Lasso
sparsity
$\lambda \uparrow \quad w \to 0$

Ridge

alpha	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6
0.0000	-9.160885	-205.462260	516.684624	340.627341	-895.543609	561.214533	153.884786	126.734316	861.121400	52.419828
0.0001	-9.118336	-205.337133	516.880570	340.556792	-883.415291	551.553259	148.578680	125.355917	856.480254	52.467627
0.0010	-8.763583	-204.321125	518.371729	339.975385	-787.690766	475.274718	106.786540	114.632063	819.739542	52.872100
0.0100	-6.401088	-198.669767	522.048548	336.348363	-383.709187	152.663678	-66.060583	75.611090	659.869402	55.828128
0.1000	6.642753	-172.242166	485.523872	314.682122	-72.939323	-80.590053	-174.466515	83.616653	484.363285	73.584154
1.0000	42.242217	-57.305508	282.170831	198.061386	14.363544	-22.551274	-136.930053	102.023193	260.104308	98.552274
10.0000	21.174004	1.659796	63.659772	48.493240	18.421492	12.875448	-38.915435	38.842464	61.612405	35.505355
100.0000	2.858979	0.629452	7.540604	5.849997	2.710879	2.142134	-4.834047	5.108223	7.448466	4.576129
1000.0000	0.295726	0.069290	0.769004	0.597829	0.282900	0.225936	-0.495607	0.527031	0.761497	0.471029
10000.0000	0.029674	0.006995	0.077054	0.059915	0.028412	0.022715	-0.049686	0.052870	0.076321	0.047241

single $x \mid y \rightarrow$

alpha	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6
0.0000	-9.160885	-205.462260	516.684624	340.627341	-895.543596	561.214523	153.884780	126.734314	861.121395	52.419828
0.0001	-9.071288	-205.337332	516.780313	340.539730	-888.652320	555.952271	150.585260	125.453044	858.639860	52.379002
0.0010	-8.264924	-204.213177	517.641106	339.751339	-826.653342	508.609613	120.899583	113.924518	836.314382	52.011583
0.0100	-1.361404	-192.944226	526.348511	332.649058	-430.205495	191.277876	-44.048113	68.990747	688.384976	47.939528
0.1000	0.000000	-113.976046	526.737112	292.635423	-82.691928	-0.000000	-152.691332	0.000000	551.077200	7.169852
1.0000	0.000000	0.000000	363.882636	27.278420	0.000000	0.000000	-0.000000	0.000000	336.135971	0.000000
10.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000
100.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000
1000.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000
10000.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000

$\rightarrow$ feature selection

$$m = \frac{\sum_{i=1}^{n} (y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n} (x_i-\bar{x})^2 + \lambda}$$

simple $x \mid y \rightarrow y = mx + b$

$\hookrightarrow b = \bar{y} - m\bar{x}$

$\bar{y} \rightarrow mean(y)$

$\bar{x} \rightarrow mean(x)$

$$m = \frac{\sum_{i=1}^{n} (y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n} (x_i-\bar{x})^2}$$

$|m| = m$

$m > 0$

$\boxed{b = \bar{y} - m\bar{x}}$ $\quad m = ?$

$< 0 \quad |m| = -m$

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \pm \lambda |m|$$

λm

$$\frac{d}{dm} \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda m$$

$$\frac{dL}{dm} = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda|m| = 2\sum (y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) + 2\lambda = 0$$

$$-2\sum \left[(y_i - \bar{y}) - m(x_i - \bar{x}) \right](x_i - \bar{x}) + 2\lambda = 0$$

$$m \sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda$$

$$m = \sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda$$

$$-2\sum\left[(y_i-\bar{y}) - m(x_i-\bar{x})\right](x_i-\bar{x})$$

$$-\sum\left[(y_i-\bar{y})(x_i-\bar{x}) - m(x_i-\bar{x})^2\right] + \lambda = 0$$

$$-\sum(y_i-\bar{y})(x_i-\bar{x}) + m\sum(x_i-\bar{x})^2 + \lambda = 0$$

$$\boxed{m = \frac{\sum(y_i-\bar{y})(x_i-\bar{x}) - \lambda}{\sum(x_i-\bar{x})^2}}$$

Lasso   coeff   sparsity

$\boxed{\text{for } m>0}$ $\lambda>0$

$$m = \frac{\boxed{\sum(y_i-\bar{y})(x_i-\bar{x})} - \lambda}{\sum(x_i-\bar{x})^2}$$

for $m=0$

$$m = \frac{\sum(y_i-\bar{y})(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

for $m<0$

$$m = \frac{\sum(y_i-\bar{y})(x_i-\bar{x}) + \lambda}{\sum(x_i-\bar{x})^2}$$

$m \rightarrow$

$\lambda \uparrow$

$$m = \frac{YX - \lambda}{X^2} \quad \left\{ \begin{array}{l} YX = 100 \\ X^2 = 50 \end{array} \right.$$

$2, \frac{9}{5}, 1, 0, -1, 5$ $0 \rightarrow$

$\lambda = 100$

$\boxed{m=0}$ 130

$$m = \frac{100-\lambda}{50}$$

$\lambda = 0$, $m = 2$ | $\lambda = 10$, $m = \boxed{\frac{9}{5}}$

$\lambda = 50$ $m = 1$

$\lambda > 100$

$m = -1$ 150

$$m = \frac{YX+\lambda}{X^2} = \frac{100+\lambda}{50} = \frac{100+150}{50} =$$

$m = 5$

λ -ve

$m<0$ $\lambda>0$

$$m = \frac{\sum(y_i-\bar{y})(x_i-\bar{x}) + \lambda}{\sum(x_i-\bar{x})^2}$$

$$m = \frac{-100-\lambda}{50} = \frac{-100-150}{50} = -5$$

$$m = \frac{-100+\lambda}{50}$$

$\lambda = 0$ $m = -2$

$\lambda = 50$ $m = -1$

$\lambda = 100$ $m = 0 \rightarrow 1$

$\lambda = 150$

$\lambda = 150$, $m = -5$

1) $0 \rightarrow$

2) stop

$\boxed{m=1}$

$$m = \frac{\overset{0.1}{\sum(y_i-\bar{y})(x_i-\bar{x})} \rightarrow 0X}{\sum(x_i-\bar{x})^2 + \lambda}$$

0

$\lambda = 10000000$

Denominator

$\lambda \rightarrow$

Ridge $<<0$

$==0$

Lasso $\rightarrow \lambda \rightarrow$ numerator

ElasticNet Regression

Saturday, June 12, 2021 1:04 PM

Ridge $\lambda(w_1^2 + w_2^2 + \dots + w_n^2)$

L_2

$$L = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \|w\|^2$$

mse

overfitting

$\lambda \uparrow \quad w \downarrow 0$

100 cols >

Lasso $\lambda(|w_1| + |w_2| + \dots + |w_n|)$

L_1

$$L = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \|w\|$$

mse

$\lambda \uparrow \quad w \rightarrow 0 \rightarrow$ feature selection

EN Reg.

$$L = \sum (y_i - \hat{y}_i)^2 + a\|w\|^2 + b\|w\|$$

$\{\lambda, l1\_ratio\}$

$\boxed{l_1, \lambda}$

$\lambda = a + b$

$l1\_ratio = \frac{a}{a+b}$

$l1 = \frac{a}{\lambda}$

$a = l_1 \times \lambda$

$b = \lambda - a$

$\lambda = 1 \quad l1\_ratio = 0.5$

$a = 0.5 \quad b = 0.5$

$l1\_ratio > 0.9$ 90% rid 10% lasso

Input cols → multicollinearity (ElasticNet)

$x_1 \mid x_2$