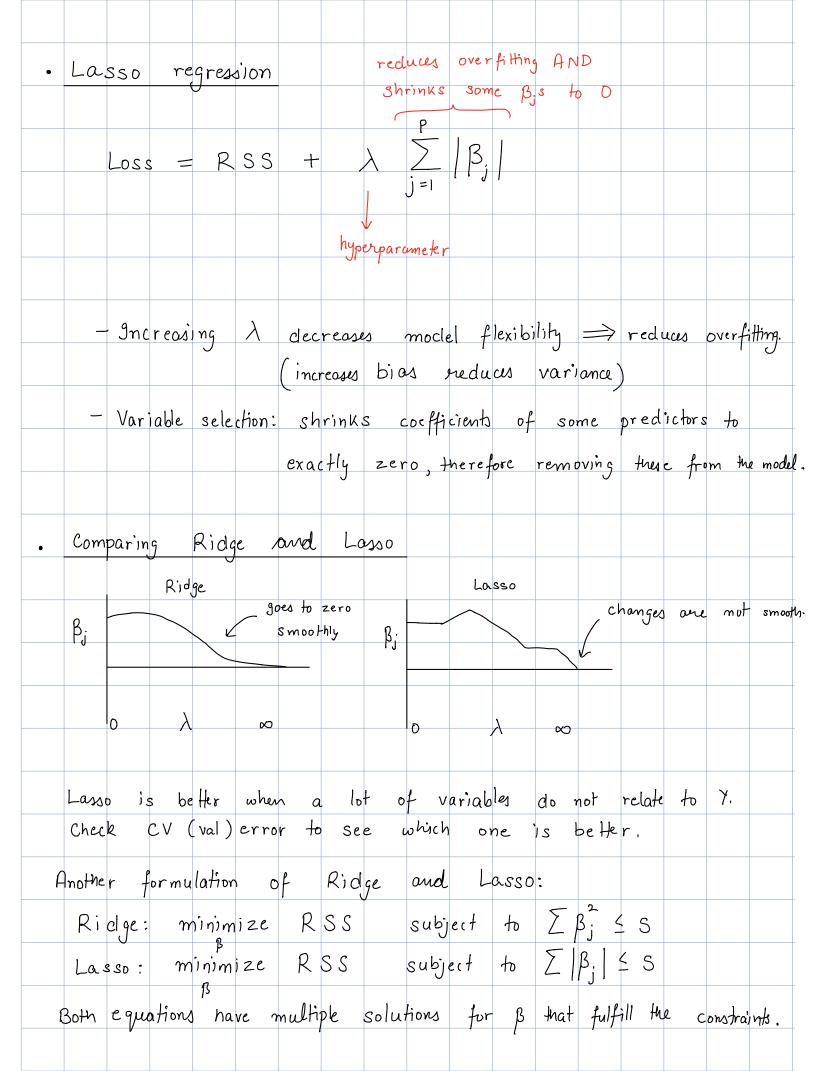
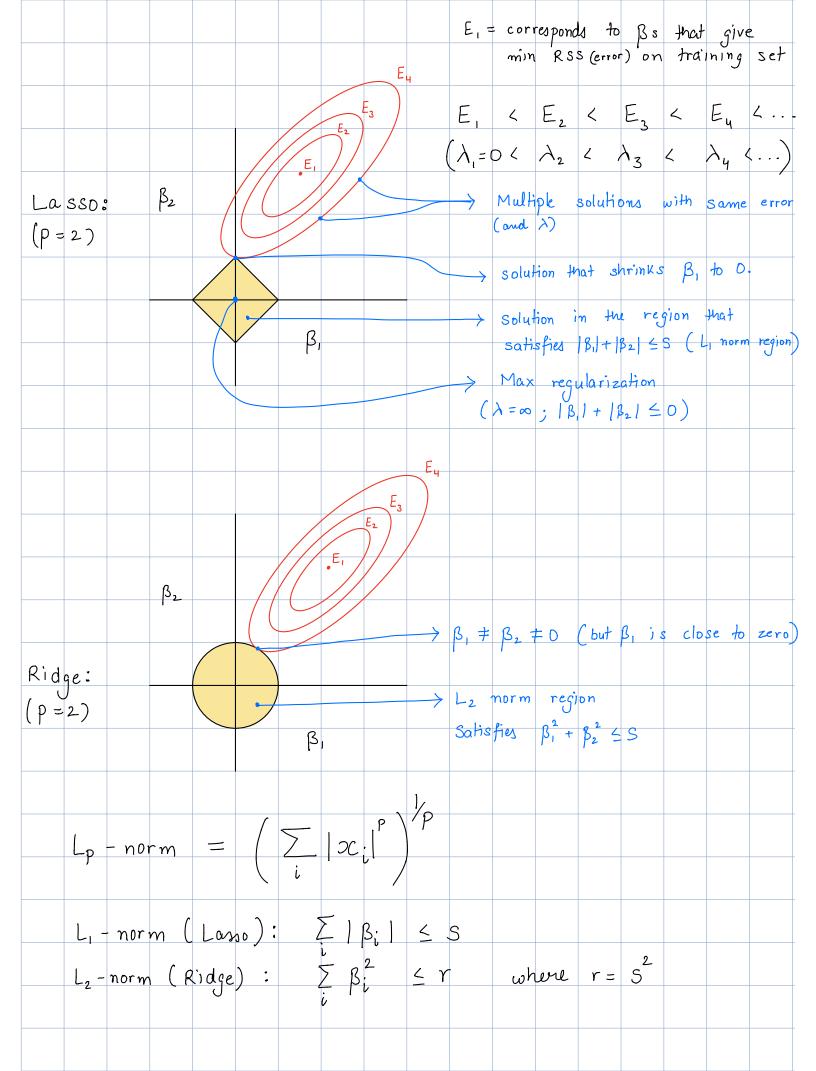
of irrelevant predictors to zero (or close to zero). How: Change the cost/loss function. $ \begin{array}{ccccccccccccccccccccccccccccccccccc$	_	4	_						_								
Ridge regression $ \begin{array}{c} \text{Ridge regression} \\ \text{rectures overfitting} \\ \text{Loss} = RSS + \lambda \sum_{j=1}^{p} \beta_{j}^{2} \\ \text{hyperparameter} \\ \\ \text{- Increasing } \lambda \text{ observations bias reduces variance} \\ \text{- All features need to be standardized because in vanilla linear regression, changing scale of a predictor (c \times x_{j}) led to automatic adjustment of the coefficient (t \times \beta_{j}) so that \beta_{j} \propto_{j} was constant. But, for ridge, the loss function also has \beta_{j}^{2} terms so \beta_{j} \propto_{j} won't be constant and will lead to bin estimate of \beta_{j} if \alpha_{j} is not standardized.$	Goo	J :	To	im	prove	· m	noclel	per	ifor mo	nce	by	red	ucin	9 0	verfi	Hing	
How: Change the cost/loss function. Ridge regression reduces overfitting Loss = RSS + λ $ \int_{j=1}^{p} \beta_{j}^{2} $ hyperparemeter - Increasing λ decreases model flexibility \Rightarrow reduces overfitting (increases bias neduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor (cxx_{j}) led to automatic adjustment of the coefficient (cxx_{j}) so that cxx_{j} was constant. But, for ridge, the loss function also has cxx_{j} for so cxx_{j} if cxx_{j} is not standardized:			To	iπ	prove	m	odel	in	erpr	etab	ility	by	rea	ducin	g C	seffic	`iev
Ridge regression $ \begin{array}{c} \text{Ridge regression} \\ \text{rectues overfitting} \\ \text{Loss} = RSS + \lambda \sum_{j=1}^{p} \beta_{j}^{2} \\ \text{hyperparemeter} \\ \\ \text{- Increasing } \lambda \text{ olecreases model flexibility} \Rightarrow \text{reduces overfith} \\ \text{(increases bias reduces variance)} \\ \text{- All features need to be standardized because in vanilla linear regression, changing scale of a predictor } \\ \text{($c \times x_{j}$) led to automatic adjustment of the coefficient} \\ \text{($c \times \beta_{j}$) so that } \beta_{j} \propto_{j} was constant. \\ \text{But, for ridge, the loss function also has } \beta_{j}^{2} \text{ terms so } \beta_{j} \propto_{j} won't be constant and will lead to bin estimate of } \beta_{j} \text{ if } \alpha_{j} \text{ is not standardized}. \\ \end{array} $			of	jr	relev	ant	pr	edic	ors	to	Zť	ero (or	close	to	zer	>)
Loss = RSS + λ $\sum_{j=1}^{p} \beta_{j}^{2}$ hyperparameter - Increasing λ decreases model flexibility \Rightarrow reduces overfith (increases bias reduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor ($c \times \lambda c_{j}$) led to automatic adjustment of the coefficient ($c \times \beta_{j}$) so that $c \times \beta_{j} \times \beta$	Hou) •	Cho	mge	H	re	cos	t / los	s f	unc t	ion.						
Loss = RSS + λ $\sum_{j=1}^{P} \beta_{j}^{2}$ hyperparameter - Increasing λ decreases model flexibility \Rightarrow reduces overfith (increases bijas treduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor ($c \times x_{j}$) led to automatic adjustment of the coefficient ($c \times \beta_{j}$) so that $c \times \beta_{j} \times \beta$	Ric	ge	re	gress	ion			reclu	(l)	ove r f	i Hina						
hyperparameter - Increasing λ decreases model flexibility \Longrightarrow reduces overfith (increases bijas reduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor (c x x;) led to automatic adjustment of the coefficient (t x \beta;) so that \beta; \alpha; was constant. But, for ridge, the loss function also has \beta; tensor so \beta; \alpha; won't be constant and will lead to bin estimate of \beta; if \alpha; is not standardized.									P								
hyperparameter - Increasing λ decreases model flexibility \Longrightarrow reduces overfith (increases bijas reduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor (c x x;) led to automatic adjustment of the coefficient (t x \beta;) so that \beta; \alpha; was constant. But, for ridge, the loss function also has \beta; tensor so \beta; \alpha; won't be constant and will lead to bin estimate of \beta; if \alpha; is not standardized.		Los	s =	= F	\ S S	3 -	_	λ	$\sum_{i=1}^{n}$	ß							
- Increasing λ decreases model flexibility \Rightarrow reduces overfith (increases bijas reduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor (Cx x;) led to automatic adjustment of the coefficient (\dot{c} × $\dot{\beta}$;) so that $\dot{\beta}$; \dot{x} ; was constant. But, for ridge, the loss function also has $\dot{\beta}$; tensor so $\dot{\beta}$; and world lead to big estimate of $\dot{\beta}$; if \dot{x} ; is not standardized.									J =1	. 0							
- Increasing λ decreases model flexibility \Rightarrow reduces overfith (increases bijas medicus variance) - All features med to be standardized because in vanilla linear regression, changing scale of a predictor (Cx x;) led to automatic adjustment of the coefficient ($\dot{z} \times \beta_{j}$) so that $\beta_{j} \propto_{j} \omega_{as}$ constant. But, for ridge, the loss function also has β_{j}^{2} tensor so $\beta_{j} \propto_{j} \omega_{as}$ is not standardized.							hu	V perpa	ramete	r							
(increases bias reduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor $(c \times x_j)$ led to automatic adjustment of the coefficient $(c \times \beta_j)$ so that $\beta_j \times_j was$ constant. But, for ridge, the loss function also has β_j^2 terms so $\beta_j \times_j was$ constant and will lead to biase estimate of β_j if α_j is not standardized.								·									
(increases bias reduces variance) - All features need to be standardized because in vanilla linear regression, changing scale of a predictor $(c \times x_j)$ led to automatic adjustment of the coefficient $(c \times \beta_j)$ so that $\beta_j \times_j was$ constant. But, for ridge, the loss function also has β_j^2 terms so $\beta_j \times_j was$ constant and will lead to biase estimate of β_j if α_j is not standardized.		_	Crea	jng	λ	de	crea	ses	moc	lel	flexil	511;Hy	==;	> rec	luces	over	fitt
- All features need to be standardized because in vanilla linear regression, changing scale of a predictor $(C \times X;)$ led to automatic adjustment of the coefficient $(\dot{C} \times \beta;)$ so that $\beta; x;$ was constant. But, for ridge, the loss function also has $\beta;^2$ tensor so $\beta; x;$ won't be constant and will lead to bis estimate of $\beta;$ if $x;$ is not standardized.				J		(in	crease	a bi	as	red	uces	Vai	ionce	2)			
vanilla linear regression, changing scale of a predictor $(c \times x_j)$ led to automatic adjustment of the coefficient $(c \times \beta_j)$ so that $\beta_j \times_j was$ constant. But, for ridge, the loss function also has β_j^2 terms so $\beta_j \times_j was$ constant and will lead to bin estimate of β_j if α_j is not standardized.																	
vanilla linear regression, changing scale of a predictor $(c \times x_j)$ led to automatic adjustment of the coefficient $(c \times \beta_j)$ so that $\beta_j \times_j was$ constant. But, for ridge, the loss function also has β_j^2 terms so $\beta_j \times_j was$ constant and will lead to bin estimate of β_j if α_j is not standardized.		- AI) _1	featu	r PA	on re	d	to	be.	st	andar	cliz	ed	hei	.OULS	Se Im	
($C \times X_j$) led to automatic adjustment of the coefficient ($C \times \beta_j$) so that $\beta_j \times \zeta_j$ was constant. But, for ridge, the loss function also has β_j^2 terms so $\beta_j \times \zeta_j$ won't be constant and will lead to bin estimate of β_j if X_j is not standardized.			,														
($\dot{z} \times \beta_{j}$) so that $\beta_{j} \propto_{j}$ was constant. But, for ridge, the loss function also has β_{j}^{2} ter so $\beta_{j} \propto_{j}$ won't be constant and will lead to bid estimate of β_{j} if α_{j} is not standardized.						_				•			V		ı		
But, for ridge, the loss function also has β_1^2 terms so $\beta_2 \propto 1$, won't be constant and will lead to bin estimate of β_3 ; if ∞ ; is not standardized.		,		- /													
so β , ∞ ; won't be constant and will lead to big estimate of β ; if ∞ ; is not standardized.			ι .	رادا) &		that		Dj 2	j	was	Co	nsta	M7 •		2	
estimate of B; if x; is not standardized.		130	ut,	708		ridge	- ر-	the	083	7	unc 17	0m	also	ho	<u>)</u>	β Ι	J eri
		<i>e</i> 2	ltima:	te	of	βĵ		f	oc;	ľ	`S	not	S	tend	ard	ized.	
									_								





Note	, , (Choo.	se	λ	by	tryi	ng (out	vari	ous	valu	us or	nd	pickin	g Hu	l 0n∢	2
		that	gi	ves	loω	est	С	V e	rror	อท	V	alid	aHon	5 e-	- .		
			U														