0	Date	=0
4		

Assignment 1 Ouestion 2 Show that ridge regression can be revertler as a non regularized linear regression min  $\frac{1}{||Xw+b1-y||_2^2 + \lambda ||w||_2^2}$   $||we||_{R^a, b \in ||R||} = \frac{1}{2n} \left( \frac{||Xw+b1-y||_2^2 + ||Xw||_2^2}{2n} \right)$ = \( \( \text{\pm} \text{\pm} + \text{\su} \text{\su} \text{\sum} \\ \text{\su} \text{\su} \\ \text{\su} \text{\su} \\ \text{\su} \text{\su} \\ \text{\su} \text{\su} \\ \  $=\frac{1}{2n}\left[\begin{array}{cccc} X\omega+b1-y \\ \hline 2n \end{array}\right]\left[\begin{array}{cccc} X\omega+b1-y \\ \hline \sqrt{2n}\lambda I J\omega \end{array}\right]\left[\begin{array}{cccc} X\omega+b1-y \\ \hline \sqrt{2n}\lambda I J\omega \end{array}\right]$  $= \frac{1}{2n} \left[ \begin{array}{c|c} X \omega + b & -\omega \\ \hline 2n & \overline{2n} & \overline{3} \omega + 0 \end{array} \right] \quad \text{and} \quad \omega \approx d \sin d$  $= \frac{1}{2n} \left[ \frac{x_w + b_1}{52n x_w} - \frac{y}{0d} \right]_2^2$   $= \frac{1}{2n} \left[ \frac{x_w + b_1}{x_w + b_1} - \frac{y}{0d} \right]_2^2$   $= \frac{1}{2n} \left[ \frac{x_w + b_1}{2n x_1 dw} - \frac{y}{0d} \right]_2^2$   $= \frac{1}{2n} \left[ \frac{x_w + b_1}{2n x_1 dw} - \frac{y}{0d} \right]_2^2$   $= \frac{1}{2n} \left[ \frac{x_w + b_1}{x_w + b_1} - \frac{y}{0d} \right]_2^2$   $= \frac{1}{2n} \left[ \frac{x_w + b_1}{x_w + b_1} - \frac{y}{0d} \right]_2^2$  $= \frac{1}{2n} \left[ \begin{array}{c|c} x & 1 & w \\ \hline - \sqrt{2n} & 1 & 0 \end{array} \right] \left[ \begin{array}{c|c} w & - & y \\ \hline - & & \end{array} \right]$ 

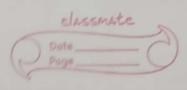
Our O du 1 1/X w + 61 - y 1/2 + 2 1/4 1/2 25+tern=1 ((Xw+b1-y) (Xw+b1-y)) =1 (wTx+11b-y) (Xw+b1-y) = 1 (w x x x w + 21b x w - y x x + w x x 61 + 67b 2n - y x b1 - w x x y - 16 y + y x y Now derive wrt w  $= \frac{1}{2n} \left( \frac{1}{2w} \left( w^{T} X^{T} X \omega + b^{T} X \omega - y^{T} X \omega + w^{T} X^{T} b - y^{T} X \omega + w^{T} X^{T} b \right)$   $= \frac{1}{2n} \left( \frac{1}{2w} \left( w^{T} X^{T} X \omega + 2 w^{T} X^{T} b - 2 w^{T} X^{T} y \right)$   $= \frac{1}{2n} \left( \frac{1}{2w} \left( w^{T} X^{T} X \omega + 2 w^{T} X^{T} b - 2 w^{T} X^{T} y \right)$  $=\frac{1}{2n}\left(\frac{\partial}{\partial w}\left(\omega^{T}X^{T}Xw+2(X^{T}b)^{T}w-2(X^{T}y)^{T}w\right)\right)$  $= \frac{1}{2n} \left( \frac{3}{3\omega} \left( \omega^{T} X^{T} X \omega \right) + 2 X^{T} b - 2 X^{T} y \right)$ Le froduct rule of matrix fortial derivative  $= \frac{1}{2b} \left( 2 X^{\mathsf{T}} X \omega + 2 X^{\mathsf{T}} b - 2 X^{\mathsf{T}} y \right)$  $= \sum_{x} X^{T}(Xw + b - y)$ 

2nd term  $\Rightarrow 22 |w||^2 = 2 |w|$ 

Clasemate Date Page

Combining terms Xw+61-9/12 (Xw+b1-y) T (Xw+b1-y) [ (wTXT+bT-y,

. Broved



Oriestion 2 3. Algorithm Submitted 4. Algorithm Submitted 5. Comparing run time of 2 afferables Eas 7 = 10 Runtine of dosed form approach = 0.0003 Runtine of gradient descent approach = 0.442 Closed form Training error =

Test error =

Test error = 0.062297 0.062297 0. 528986 λ=10 Closed form Teraining larger = 0.41589 Teraining lass = 0.45481 Test error = 0.44393 λ=0 (-1-+0.+ To in larger = 0.06259 0 - 44393 Gradient Descent Training laron = 0.062595

Teraining loss = 0.062595

Test error = 0.050952 7 = 10 Goodiert Resent Tensining grows = 0.41642 Testing loss = 0.45481 Testing error = 0.44419 I believe  $\chi = 10$ , closed form is a better opproach lecause it not only has a faster runtime but also a smaller testing error ... For this dataset,  $\chi = 10$  closed form approxis leetter.