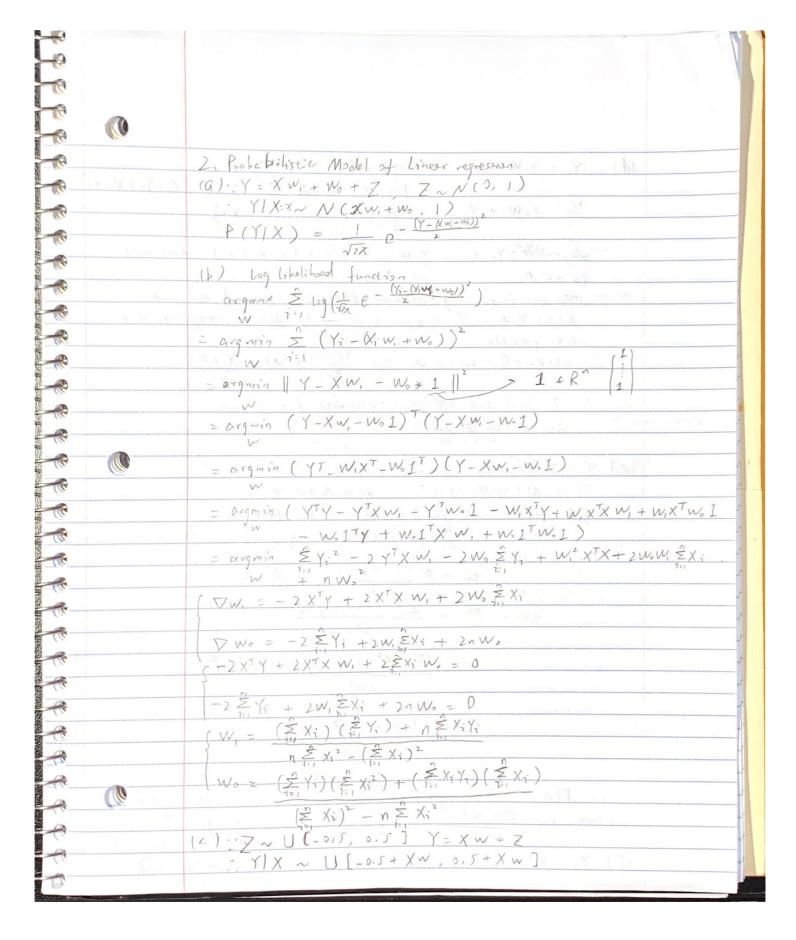
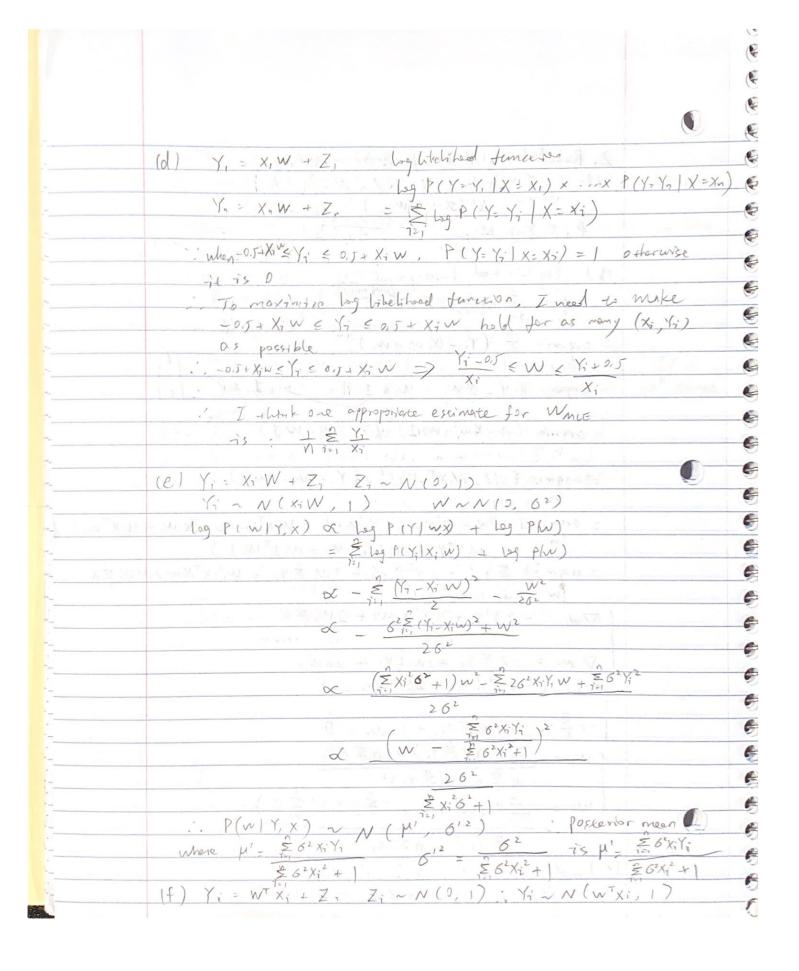
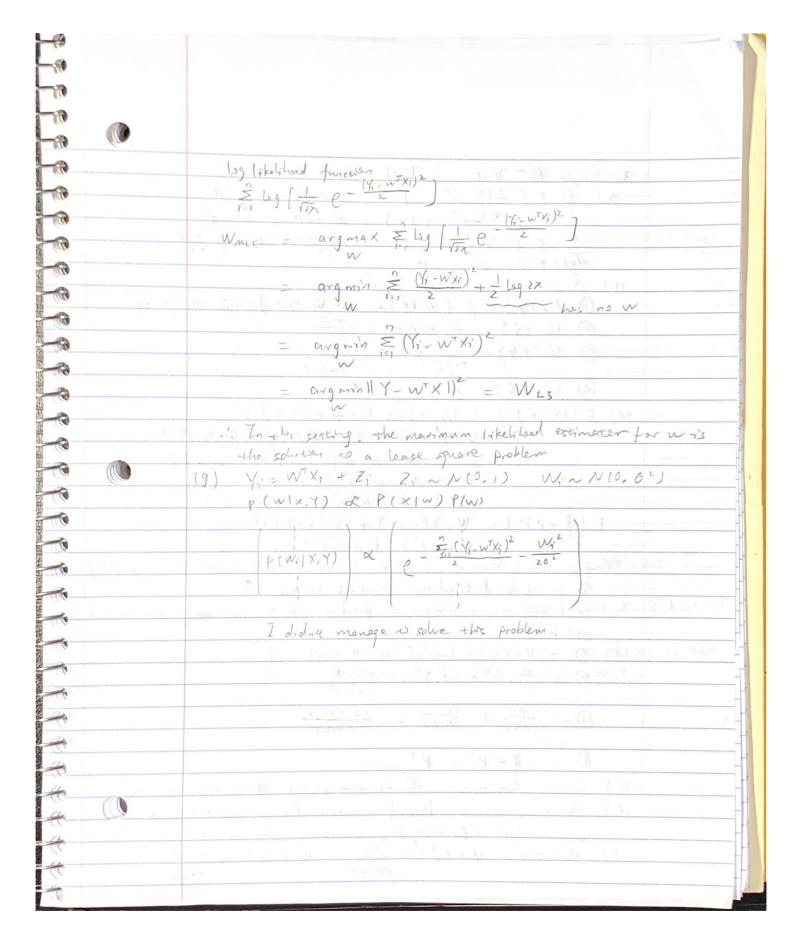
N TO		
-8		
	. (()	CS189 HW3
-0	March CA	(a) in Appellance application of the second
	1 laboure	0* = a-gricx = log P(Xil 0) + n log P(0)
	1000 000	= largmin 11 X0 - Y 1/2 + 62 11 0 - 40 1/2
		Vo (11×0-×11/2+ 002 110-101/2)
		$= \frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \right) + \frac{1}{2} \left(\frac{1}{2} \left($
40		$= V_{\mathfrak{d}} \left((\times \mathfrak{d})' \times \mathfrak{d} - (\times \mathfrak{d})' \times - (\times \mathfrak{d})' \times - (\times \mathfrak{d})' \times + (\times \mathfrak{d})'$
	7.14	$=2X^{T}\times 0-2X^{T}y+\frac{26^{2}}{6h}(\theta-\mu_{\theta})=0$
	100 2 301	$X^{T} \times 9^{*} - X^{T} + 6^{*} (8^{*} - 10) = 0$
		$D^* = (X^T X + \frac{6}{6k}I)^T (X^T Y + \frac{6}{6k} \mu_{\theta})$
		(b) The two digues for $\delta_h = 1$ and $\delta_h = 10$ one included in Appendix
		The code is in problem 1 - starter. py from # Generating Pata and
	Code also	Sh to 10. The range and variation of prior obviously increases
	fincludes the im	This is because on is the secondard deviation of prior distribution
	of MLE, MAP	On increases means the variance of prior increases.
	and privary	(6) The two figures for to = () and () are included in Appendix =
-		The code is in problem 1 - starter py
-		When change to (), the center of protor plot moves from (0,0)
		to (-1, 5) x when change to [3], the center of prior plat moves
0		from (-3, 5) to (5, 3). Unit doesn't change according -
0		from (-5, -t) to (3, 3). O MLE closesnet change according to the plots with respect to prior (d) OMAP = argmax \(\frac{9}{1-1} \) bg \(P(\times 10) \) + \(\times \(P(0) \) \) (\(\frac{1}{2} \) \(1
0		(A) AMAP = (1) (Y ₁ -X ₁ 0 ₁) ² (9-µ)
C		= arg max & log (\frac{1}{\tan \text{0}} \text{0} \text{1}
0		
	(0	$= \frac{\alpha rgmin}{\delta \epsilon R^{d}} = \frac{(\gamma_{i} - \gamma_{i} \theta_{i})^{2}}{26^{\circ}} + \frac{(\theta - \mu)_{i}}{\delta h}$
4		$= \underset{\text{argmin}}{\operatorname{argmin}} \frac{\sum_{i=1}^{n} (y_i - y_i \theta_i)^2 + \frac{26^2}{6h} \theta - \mu }{\frac{26^2}{6h} \theta - \mu }$ $= \underset{\text{argmin}}{\operatorname{argmin}} y - x \theta ^2 + \frac{26^2}{6h} \theta - \mu $
4		= argmin y - X 0 2 + 26 0 - 1
1		O'ER'







```
0
                                                                             0=
3. Simple
            Bing - Varione Trades/
                                                                            0=
                  X1+X1+..+Xn
                  = 1 ( Var (X1) + -- 1 Var (X-)
                  = (171)2
                                      (n+100)2
        Var (12) = 0
(6) E[(\hat{X}-X')^2] = V_{ar}(\hat{X}-X') + E(\hat{X}-X')
                     = Var (X) + Var (X') + {E((X-1)+(1-X'))}2
                         Ver12)+62+[E(X-Y)+E(1-X')]2
                     = Var(x)+ 02+ [E(x-1)]2
                     = Var (2) + 62 + [bios (2)]2
                  = Var (2-4) + LE (2-4) 72
                  = Var (x) + (bias(x))
                there is a 62 in E [(X-X')2] but not in E[Q-M2]
     & is calculated based on training samples
    so I will perform worse or fresh sample X' therefore the error
    will be larger.
                                 n62+12=
   0 = Ver(x) + bias (x)
                                  (n+1)2
       = (nin) +
                      10 H2 =
                                  (U+NO)2
             0 + W2 =
                           1) is when No = 1 @ is when no = 00
     As No increases, bias increases (since bias
   orevelly bias is decreasing) variance decreases

N_s = dn, V_{ar}(\vec{X}) = \frac{n6^2}{(n+dn)^2} bias (\vec{X}) = \frac{d}{n}
                                                       anH
                                                       n + dn
```

STATE OF THE PROPERTY OF THE P

expected total error = not (an H) (n+xn) (n+ dn)2 3 d expected cold error = nº42 × 2 d (n+ an)2 - (no2+2 nº42) x) (n+do)xn 2 n2 M2 d(n+ dn)2 = 2 n (0+ dn) (n 62 + d2n2 M2) 1/2 (n+2n) d = n82 + x2n2 42 n' m' x + n' m' x = n' m' x - + n 62 $\alpha = \frac{100}{100} = \frac{100}{100}$ $\alpha = \frac{100}{100} = \frac{100}{100}$ (h) When o is large and plis small relose to 0) (1) X' = (X-1/2 more day of day of E(X) = E(X-H) = H-H. Var (X') = Var (X-1) = Var (X) = 0= In vidge regression, as & thereeses, model brias T model variance & This is very similar to & s. influence of vertence. So in ridge regression, we can use cross valor the I volue with the smallest validacian error 1

	4. Pebotic Learning of Controls from Demonstration and Trages.
	(a) The Oth 10th and 20th images in the training set are
	platted in the appendix
	Their corresponding control vectors are shown in appendix
	The coole is in robotic ridge code starter. Py under section
	# 4(a).
	(b) The code to clothes is under section # 4(b)
_	When I attempt to do this, it will raise Singular matrix error
_	This is because det (XTX) = 0, XTX is not invertible
	(C) For each A in [0,1, 1,0, 10,0, 100,0, 1000.0] the recult of
	shown in appendix
	The code is under section # 4(C)
	(d) The result is shown in appendix
	The code is under section # 4(d)
	(e) The result is shown in appendix
	The code is under section # 4(e)
-	By increasing I value the breas will increase because
	It means the model complexity will decrease, therefore hims
100	will therease.
	By increasing I value, the varience will decrease because
	It model complexity I, therefore varience V
	(f) The regule is shown in appendix
	The ode is under section # 4(f)
	Y to visit the last control to the c