Parcial HA I sea el modelo de regresion to= d(xn) without con € tn ∈ R, xn ∈ R° 3 =1, w ∈ R9, Φ: R° -> R9, Φ ≥ P, y no ~ N (no 10, 00) - minimos cuadrados modelo 6n= 0 (xn) wt nn w-> payametros a estimar O(xn) -> transformación de las variables de entrada nn -7 ruido sussiano media o varianza constante 3(w) = S CEn - Ø Cxn) Tw)2 vector de predicción Ow. t = [t1, t2 -- tN] & RN vector de errores 6- Dw 6 = (x2) | GRNX9 Función de costo Ø (XN) 30w)=11t- 0w112 WERG Derivamos Funcion de costo Jan) = (t- ou) (t- ou) = 6 t- 2t out w o ou se derive con respecto a w 3] = -2 9 t + 2 0 T ou -2 \$ t+ 2 \$ T &w =0

.

3

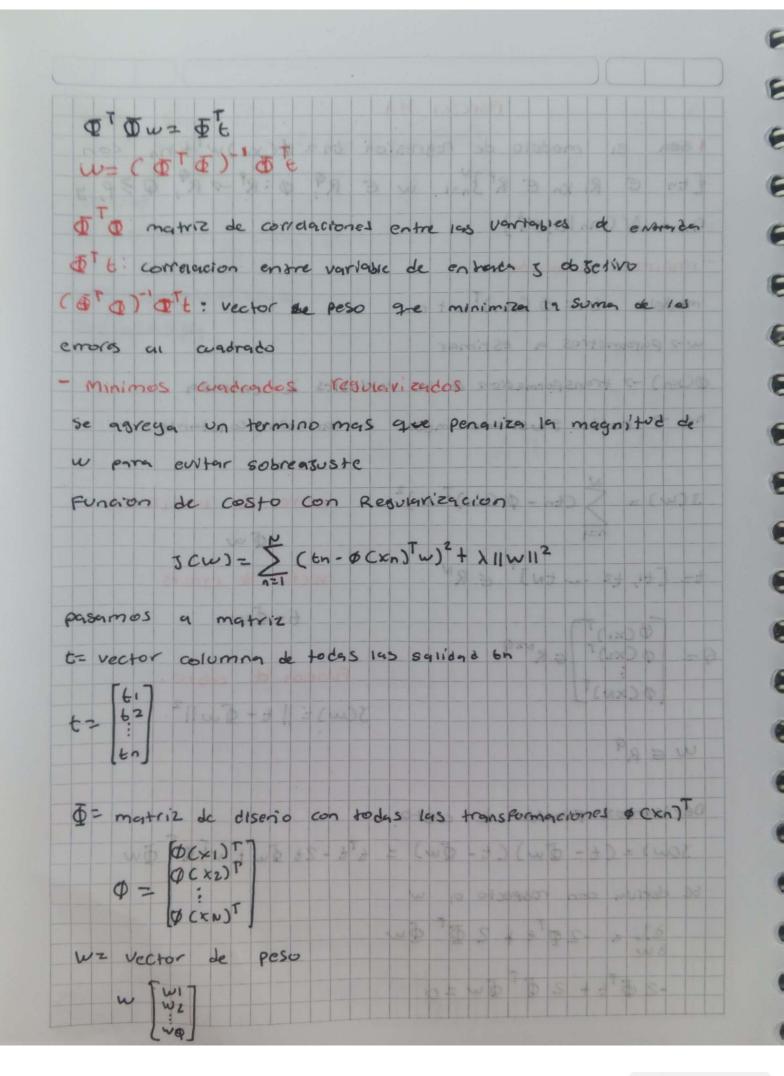
3

3

3

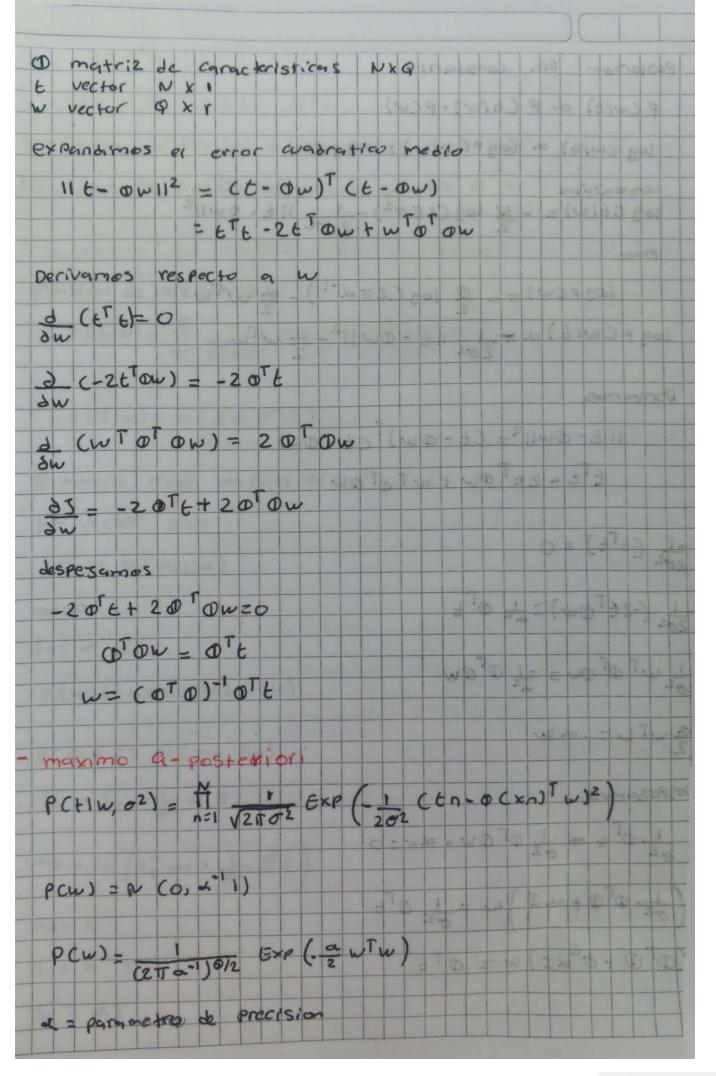
3

3



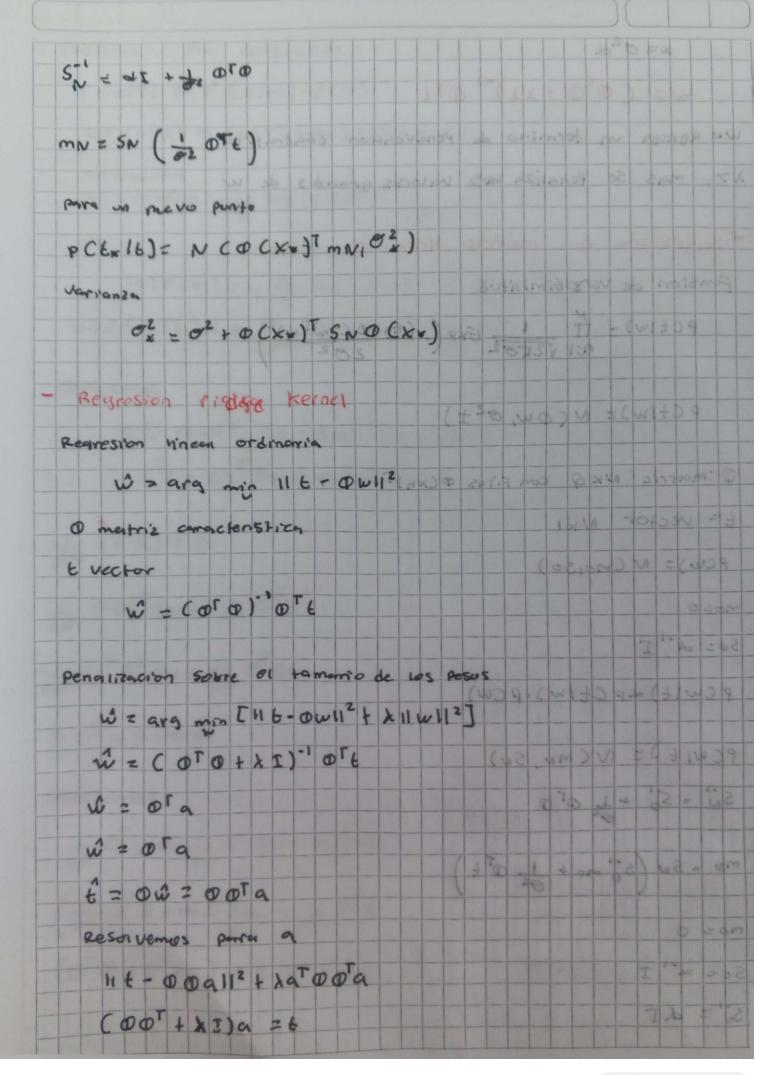
E = dw valor preducno para todos los datos 0 * w = 0 (xn) w 114112 = \ \ a;2 11t - 0w112 = 2 Ctn - 0 Cxn) [w)2 11 W112 = WTW * x -> x 11 W112 = kWTW 3cm) = 11 t - Qu112 + x 11w112 Q GRNX9 matriz de diseño t E RM vector objetivo X20 parametro regularización controla el trade off entre complesione del modero 3 asuste expandimos la Función de costo Jan = (t- ow) T(t- ow) + xwTw = 6 L= 2t Ow + wTOTOW primer termino Segundo - + AWTW termino Funcion completa Jan = tte-2t on + wto on + twin Derivamos respecto a w dernada de E respecto a w es o derivada de 25 du es -20 TE derivada de wroto w es 20 Tow

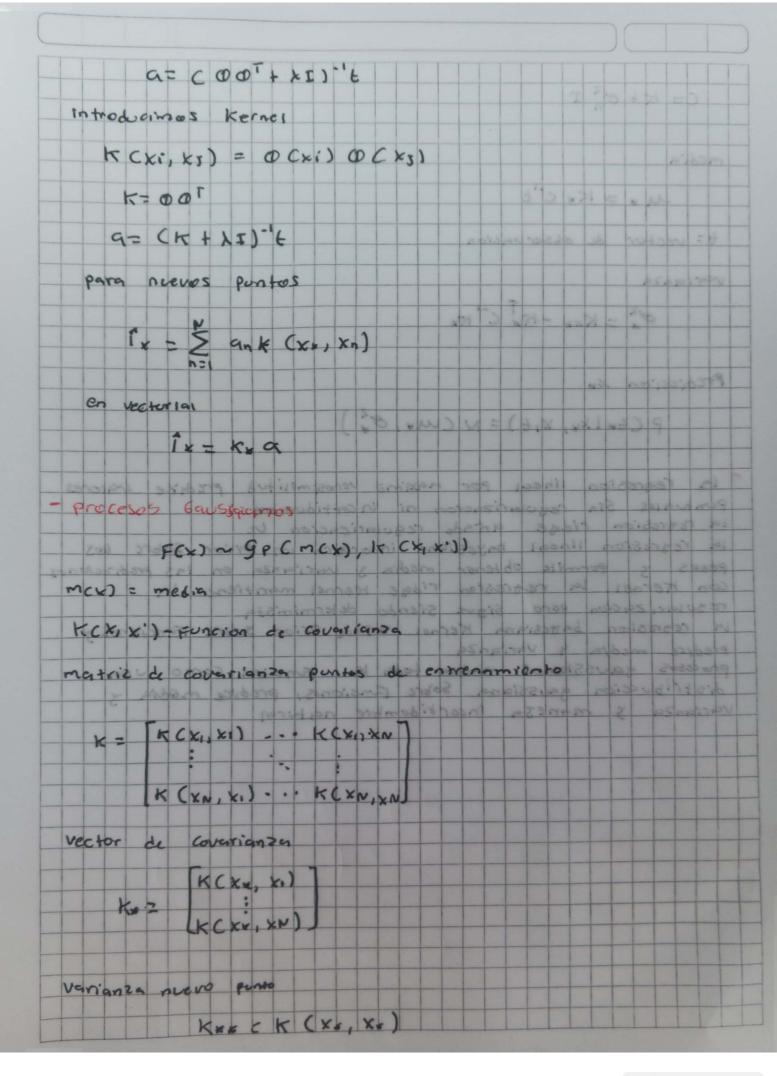
33 = -2 0 t + 2 0 pw + 2 x w nos daria 3 x = -2 0 E + 20 Tow + 2 xw despesamos -20 E +2 C OT O + X I) W = 0 (OTO +XI) w = OTE w= coto + xI)-1 ort 1) 1 matriz de correjación entre las voriables de entrada At termino de regularización, controla la magnitud del coentrola mayor la penalización > Se reduce les magnifices de - maxima verosimilitud o runcion de verosimilitud P(61 w, 02) = IT 1 GMP (-202 (En -0 CXn) Tw)2) Pasamos a wag logp (61 w, 02) = - N log (2402) -1 5 (6n - 0 (xn) w)2 5 c w) = 5 (tn-0 cxn) Tw)2 5 CEn-OCXNJT WJ2 = 11 t-6 W113



Posterior Sin constante PCUIE) & P(E/W).PCW) log curt) a loge(tlw) + log pcw) Verosimilitud 109 CEIW) z - N 109 (2502) - 1 116-0W112 109 PCW) = _ @ 109 (211 d 1) - = w Tw 109 P CWIE) a = 1 116 - OW 112 - 2 WTW Ochvamos 11 E- Ou113 = CE-OW) (+ OW) E'E-26 OW + WTOTOW =1 (tre) = 0 -1 (-2+ Tow) = 12 OTE -1 w T O T OW = -12 O TOW · de wiw = - du Despesamos. 1 0 6 - 1 0 0 0 w - aw = 0 (1 0 0 + 2 I) w = 1 0 E (0 0+02 x =) w = 0 TE

W= (OTO + XI) OTE we agrega un termino de penguencion contratica 27, mas se penalize 165 valores grandes de W Bayesiano con modelo lineal gas spiano Funsion de verosimilità PCtIW) - 17 1 Exp (- (En-0 Cxn) [w)2) PC+IND= N(OW, 02 ±) 1 matric NXQ con Fires O(xn) To - 311 E= vector NXI PCW) = N (mo,50) mozo 50= d I pcwit) = pctiw).pcw) PCWIEDE NCMW, SW) SN = 50 + 1 000 mp = SN (50 mo + 1 0 t) mo= 0 502 0 I 50 = DI





C= K+02 I media u = Kret t= vector de observacion verianza 02 = Kx - KT C Kx Prediction Xx P(6x1xx, x, 6) = N(mx, 02) la regression lineal por maximo veros; militud predice valores puntuales Sin regularization ni incortidumbre la regression ridge anade regularización la la regresion lineal bayesiana incorpora un prior sobre les pesos y permite obtener media y varianza en las predictiones con Kernel. la regression ridge Kernel memtiene la resularización pero sigue siendo determinista in regression bayesiana Kernel arrade Prior sobre Funciones y preduce media 4 varianza procesos gaussianos modera toda la Función como una distribucción gaussiana sobre funciones, preduce media y varianza y maneza incertidumbre natural