CAP 6610 Homework 3

Siddhant Mittal (6061-8545)

1. (a) Given:

$$P(y/x; \theta) \sim N(\phi^T \theta, \sigma^2)$$

 $P(\theta) \sim N(0, \sigma_o^2 I)$

From this we can deduce this:

$$P(y_i/x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\phi^T \theta - y_i)^2}{2\sigma^2}}$$
$$P(\theta) = \frac{1}{\sqrt{2\pi\sigma_o^2 I}} e^{-\frac{(\theta)^2}{2\sigma_o^2 I}}$$

Substituting the above two values in given explicit MAP formulation $minimize \sum_{i=1}^{n} -\log p(y_i/x_i, \theta) - \log p(\theta)$ we get the following:

$$minimize \sum_{i=1}^{n} -\log \frac{1}{\sqrt{2\pi\sigma^{2}}} e^{-\frac{(\phi^{T}\theta-y_{i})^{2}}{2\sigma^{2}}} -\log \frac{1}{\sqrt{2\pi\sigma_{0}^{2}I}} e^{-\frac{(\theta)^{2}}{2\sigma_{0}^{2}I}}$$

Since MAP depends on θ we remove the constant terms and we get:

$$minimize \sum_{i=1}^{n} (\phi^{T} \theta - y_i)^2 + \frac{\sigma^2}{\sigma_o^2 I} ||\theta||^2$$

Which is of the form $L(\theta) + \lambda r(\theta)$ where

$$\lambda = \frac{\sigma^2}{\sigma_0^2 I}$$

(b) Given:

$$P(y/x;\theta) \sim N(\phi^T \theta, \sigma^2)$$

$$P(\theta) = \prod_{j=1}^{m} \frac{1}{2a} exp^{-\frac{|\theta_j|}{a}}$$
$$-\log p(\theta) = -\log(\prod_{j=1}^{m} \frac{1}{2a} exp^{-\frac{|\theta_j|}{a}})$$

$$-\log p(\theta) = m \log 2a + \sum_{j=1}^{m} \frac{|\theta_j|}{a}$$

Taking from part (a) $P(y_i/x_i;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(\phi^T\theta-y_i)^2}{2\sigma^2}}$ and substituting in MAP formulation we get

$$minimize \sum_{i=1}^{n} (\phi^{T}\theta - y_i)^2 + \frac{2\sigma^2}{a} \sum_{j=1}^{m} |\theta_j|$$

$$minimize \sum_{i=1}^{n} (\phi^{T}\theta - y_i)^2 + \frac{2\sigma^2}{a} ||\theta_j||_1$$

Which is of the form $L(\theta) + \lambda r(\theta)$ where

$$\lambda = \frac{2\sigma^2}{a}$$

(c) Given:

 $p(y/x, \theta) = Pr[yu \ge 0]$ where $p(u/x, \theta) \sim N(\phi^T \theta, \sigma^2)$ We can do the following:

$$p(y/x,\theta) = Pr[yu \ge 0]$$

$$p(y/x,\theta) = \frac{1}{\sqrt{2\pi}} \int_0^\infty exp(-\frac{(u - y\phi^T\theta)^2}{2}) du$$

$$p(y/x,\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y\phi^T\theta} exp(-\frac{u^2}{2}) du$$

$$p(y/x,\theta) = \Phi(y\phi^T\theta)$$

Using part a we get

$$P(\theta) = \frac{1}{\sqrt{2\pi\sigma_o^2 I}} e^{-\frac{(\theta)^2}{2\sigma_o^2 I}}$$

Substituting in MAP formula we get

$$minimize - \log(\Phi(y\phi^T\theta)) + \frac{1}{2\sigma_0^2 I}||\theta_j||_2^2$$

Which is of the form $L(\theta) + \lambda r(\theta)$ where

$$\lambda = \frac{1}{2\sigma_0^2 I}$$

(d) Given:

$$p(y_i, x_i; \theta) = \frac{e^{-y_i \phi^T \theta}}{1 + e^{-y_i \phi^T \theta}}$$

and taking from part (b) the value of $p(\theta)$ we can do following changes in our MAP formulation and ignoring constant terms we get:

$$minimize \sum_{i=1}^{n} y_i \phi^T \theta + \frac{||\theta||_1}{a}$$

Which is of the form $L(\theta) + \lambda r(\theta)$ where

$$\lambda = \frac{1}{a}$$

2. (a) We know that

$$Prox_f(\theta) = argminf(\theta) + \frac{1}{2}||\theta - \theta_1||^2$$

To find proximal mapping for θ we need to minimize the above equation that means we can write

$$\theta = Prox_f(\theta_1)$$

Differentiating RHS and using sum rule of sub differentiation we get

$$\frac{\partial}{\partial t}[f(\theta) + \frac{1}{2}||\theta - \theta_1||^2] \in 0$$

$$\partial f(\theta) + \frac{1}{2}2(\theta - \theta_1)(1 - 0) \in 0$$
$$\theta_1 - \theta \in \partial f(\theta)$$

Since $\theta = Prox_f(\theta_1)$ we put that in

$$\theta_1 - Prox_f(\theta_1) \in \partial(Prox_f(\theta_1))$$

(b) Given $g_1 \in \partial f(\theta_1)$ and $g_2 \in \partial f(\theta_2)$ we have to prove $(g_1 - g_2)^T (\theta_1 - \theta_2) \ge 0$. By defination of subgradient of $f(\theta_1)$ and $f(\theta_1)$ we have following

$$f(\theta) \ge f(\theta_1) + g_1^T(\theta - \theta_1) \forall \theta$$

$$f(\theta) \ge f(\theta_2) + g_2^T(\theta - \theta_2) \forall \theta$$

Since it it for all θ this can also be written as

$$f(\theta_2) \ge f(\theta_1) + g_1^T(\theta_2 - \theta_1) \forall \theta$$

$$f(\theta_1) \ge f(\theta_2) + g_2^T(\theta_1 - \theta_2) \forall \theta$$

Adding above two equation we get

$$g_1^T(\theta_2 - \theta_1) + g_2^T(\theta_1 - \theta_2) \le 0$$

Simplifying gives

$$(g_1 - g_2)^T (\theta_1 - \theta_2) \ge 0$$

(c)

From part (a) we have $\theta_1 - Prox_f(\theta_1) \in \partial(Prox_f(\theta_1))$ and from part (b) we have $(g_1 - g_2)^T(\theta_1 - \theta_2) \geq 0$. Putting values of θ_1 and θ_2 in form of part (a) in part (b) equation we get

Putting

$$g_1 = \theta_1 - Prox_f(\theta_1)$$

$$g_2 = \theta_2 - Prox_f(\theta_2)$$

$$(\theta_1 - Prox_f(\theta_1) - (\theta_2 - Prox_f(\theta_2))T(\theta_1 - \theta_2) \ge 0$$

By simplyfying this we get

$$(Prox_f(\theta_1) - Prox_f(\theta_2))^T(\theta_1 - \theta_2) \ge ||Prox_f(\theta_1) - Prox_f(\theta_2)||^2)$$

(d)

By applying Cauchy-Schwartz inequality property $a^Tb \geq ||a|| ||b||$ is $b \geq ||a||$

We can modify the part equation to get nonexpansiveness property.

$$(\theta_1 - \theta_2) \ge ||Prox_f(\theta_1) - Prox_f(\theta_2)||$$

3. (a) Following algorithm is derived for solving this problem.

Given objective function. minimize I 2 max (n; To_n; Toy; + y; te) — Randomly choos select a pai. — Randomly choos select a pai. — Ind the greadout al mi by differentiation equation (1) — Fees. a conversionating to convert dass. Toy; h; = (21(QTn; -0y; 1; + 400 letta)n; . y fy; delto have is dry — new of theto) is for pronimal operation of the of a conversion of the first pronimal operation of the of a conversion of the first pronimal operation of the of a conversion of the first pronimal operation of the of a conversion of the first pronimal operation of the of a conversion of the first pronimal operation of the of a conversion of the first pronimal operation of the operation o	
minimize 1 2 max (n;To,-n;Toy; + lyite) — () - O; I. Ok	- These steps are repealed to tudition
minimize 1 2 max (n;To,-n,Toy; 1 yite) — () - O; I. Ok Miss character of milk of the surpression of the sur	Given objective function.
Randomly choc select a pi Tind the gracific al ni by differentiation equation (1) Then . O . convers porching to towered dats. Toy, h; = \left(\frac{2}{2}\left(\omega_1^{\gamma_1}\left) - \omega_2^{\gamma_1}\left(\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft)\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\ga	The same of the sa
Randomly choc select a pi Tind the gracific al ni by differentiation equation (1) Then . O . convers porching to towered dats. Toy, h; = \left(\frac{2}{2}\left(\omega_1^{\gamma_1}\left) - \omega_2^{\gamma_1}\left(\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft)\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\ga	minumise + 2 max (n; 10, - n; 10y; + y; +c) -
Randomly choc select a pi Tind the gracific al ni by differentiation equation (1) Then . O . convers porching to towered dats. Toy, h; = \left(\frac{2}{2}\left(\omega_1^{\gamma_1}\left) - \omega_2^{\gamma_1}\left(\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\reft)\reft)\reft)\reft(-\omega_1^{\gamma_1}\left(-\omega_1^{\gamma_1}\left(-\omega_1^{\ga	The state of the s
→ Randomly that select a Mi. → Find the sugrection of al Mi by differentiation equation (1) → Eas . O . Convers ponding to torough dats. √ Oy.L.: = {21 (Q, 71; - Oy. 1; + & letta) M; . y + y, O corresponding to incorrect class. √ Oy.L.: = 1 (Oy. 71; - Oy; M; + delta) M; . delto have is one → New O (theta) is four pronimal operation of the total of the tota	151:0 Solotani trach itropion . 2101
Tind the sugrection of mi by differentiation equation (1) equation (2) Toy.L.: = \(\frac{21}{21} \) (\overline{\text{G}}\), \(\frac{7}{2}; -\overline{\text{G}}\), \(\frac{7}{2}; +\overline{\text{B}}\) lette \(\frac{7}{2}; -\overline{\text{G}}\), \(\frac{7}{2}; -\text{	
Toy.h; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Ta}; - \text{Oy}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; = \(\lambda \) \(\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; \(\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; \text{N}; + \text{delta} \) \(\text{N}; \) \[\text{Toy.h}; \text{N}; + \text{delta} \) \[\text{Toy.h}; \text{N}; \text{Toy.h}; \text{N}; \\ \text{Toy.h}; \(\text{N}; \text{Toy.h}; \\ \te	- Pandonili chas selast a Mi
equation (4) FEED . O . CONDICE porching to conviced dass. Toy.h; = \left(\frac{21}{y^4y_i}, \tag{\text{W}}, \text{1; + \text{W}} \text{ketha}\right)\eta_i. O convices ponding to inconvert class. Doj.l; = \left(\text{Oj}\frac{1}{x_i} - \text{Oy}_i \text{N}_i + \text{cletta}\right)\eta_i. delto here is one There is one New O(theta) is for pronimal operation of the text of te	U
equation (1) FEED . O . CONCRETE porching to convered dass. Toy.h; = \left(\frac{21}{y^4y_i}\) O converse ponding to inconvert dass. Doj.L; = \left(\frac{0}{y^4y_i}\) delto here is one There is one of the order of pronimal operation of the of the order of t	-> find the gradient at Mi by differentiation
→ Eas . O . converponding to torveel dars. Toy!: = (21 (@TA; - Oy, 1; + & letta) A; . O coveres ponding to incorrect dars. Do;! = 1 (Ojt; - Oy; N; + delta) A; . delto here is one → new O(theta) is for pronimal operation O t+! L Ot - > (Sub qradient)	equation (1)
Doy.Li = {\frac{\frac{21}{\text{(Q\sqrt{1}; -0\sqrt{1}; +\text{\text{Ast Retta}})\end{i}}}{\text{O}\text{convert class.}} O convers ponding to inconvert class. \[\text{Doy;Li} = \frac{1}{\text{(Q\sqrt{1}; -\text{Oyi}\empty; \text{1}; + delta)}\empty.}\] \[\text{delto here is \text{One}}\] \[\text{Ot - \text{(Sub-quadient)}}\]	
Doy.Li = {\frac{\frac{21}{\text{(Q\sqrt{1}; -0\sqrt{1}; +\text{\text{Ast Retta}})\end{i}}}{\text{O}\text{convert class.}} O convers ponding to inconvert class. \[\text{Doy;Li} = \frac{1}{\text{(Q\sqrt{1}; -\text{Oyi}\empty; \text{1}; + delta)}\empty.}\] \[\text{delto here is \text{One}}\] \[\text{Ot - \text{(Sub-quadient)}}\]	- Few . O . conversion ding to convers dass.
O corresponding to incorrect class. Dojli = 1 (Ojta; - Oy; x; t delta) n;. delto here is one — new O(theta) is for pronimal operation Other ot - > (Subgradient)	
O corresponding to incorrect class. Dojli = 1 (Ojta; - Oy; x; t delta) n;. delto here is one — new O(theta) is for pronimal operation Other ot - > (Subgradient)	1/0 y.h; = 1/21 (0) 1, -0 y.1, + 400 xeen 1 1, 1
Døjli = 1 (Øjta; - Øyj x; t delta) mj. delto here is dne new O(theta) is for pronimal operation ottle ot - >(Subgradient)	The second secon
Døjli = 1 (Øjta; - Øyj x; t delta) mj. delto here is dne new O(theta) is for pronimal operation ottle ot - >(Subgradient)	O coveres ponding to incorrect dals.
dekto here is d'es d'es pronimal e purah'a 0 t+1 t ot - 2(Subgradient)	,
→ new · O (theta) is for pronimal o prah'a 0 ttl. L ot - > (Subgradient)	Joj Li = I (b) h; - Gy; n; + delta) m;
→ new · O (theta) is for pronimal o prah'a 0 ttl. L ot - > (Subgradient)	della here is DNO
0 trl. L 0 t - 2 (Sub gradient)	
0 trl. L 0 t - 2 (Sub gradient)	> new O(theta) is for pronimal operation
-> New weight that is there is CS Scanned with the lot extend from (0t - y(t)) The Camscanner	O - O Subgradient
CS Scanned with the (Oten to the y (f) The Camscanner	-> Num mai ma :Al . A in H A.
CS Scanned with the Company (0 - y AVI)	1 / wat is just y
Cambeanner	CS Scanned with the Cot CA TOO LAND
	Cambeanner

-> These steps are repeated 106 it mation

for :4 Lamba Value \$10,1,0.01g

-, y(t) is the step size which is diminishing

step size at each iteration, step

cs ight with 1 (O iteration is the start)

cs ight with 1

Proximal operator calculation:

Prox of $L_{2,1}$ Mixed Norm

The problem is given by:

$$\arg\min_{X} \frac{1}{2} \|X - Y\|_{F}^{2} + \lambda \|X\|_{2,1}$$

Where $X, Y \in \mathbb{R}^{m \times n}$.

Again, this can be decomposed into working on each column of X separately:

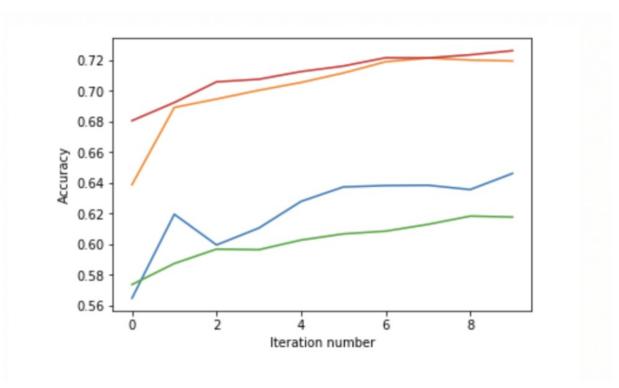
$$\begin{split} \arg\min_{X} \frac{1}{2} \|X - Y\|_{F}^{2} + \lambda \|X\|_{2,1} &= \arg\min_{X} \sum_{i} \frac{1}{2} \|X_{:,i} - Y_{:,i}\|_{2}^{2} + \sum_{i} \lambda \|X_{:,i}\|_{2}^{2} \\ &= \arg\min_{X} \left(\frac{1}{2} \|X_{:,1} - Y_{:,1}\|_{2}^{2} + \lambda \|X_{:,1}\|_{2}^{2} \right) \\ &+ \left(\frac{1}{2} \|X_{:,2} - Y_{:,2}\|_{2}^{2} + \lambda \|X_{:,2}\|_{2}^{2} \right) \\ &+ \cdots \\ &+ \left(\frac{1}{2} \|X_{:,n} - Y_{:,n}\|_{2}^{2} + \lambda \|X_{:,n}\|_{2}^{2} \right) \end{split}$$

Each term in the brackets is independent Prox Function of L_2 Norm. Hence the solution is given by:

$$\hat{X} = \arg\min_{X} \frac{1}{2} ||X - Y||_F^2 + \lambda ||X||_{2,1}$$

Where
$$\hat{X}_{:,i} = Y_{:,i} \left(1 - \frac{\lambda}{\max(\|Y_{:,i}\|_2, \lambda)} \right)$$

(c) Red line is lamba = 0.01, Green line is lamba = 0.1, Orange line is



(d) Yes it true that a large leads to a more sparse solution. Below shows exactly the same.

Weights are the order of lamba 10,1,0,1,0.01 respectively

