1. LASSO vs Ridge Homework 8

(a) $\|Xw-y\|_2^2 + \lambda\|w\|_1$

$\uparrow$

$\langle Xw-y, Xw-y\rangle + \lambda\sum|w_i|$

$= \langle Xw, Xw\rangle - 2\langle Xw, y\rangle + \langle y,y\rangle + \lambda\sum|w_i|$

$= w^TX^TXw - 2y^TXw + y^Ty + \lambda\sum|w_i|$

$= nw^Tw \quad \cdot \cdot \cdot$

$= n\sum w_i^2 - 2y^T\sum \vec{x}_i w_i + \sum y_i^2 + \lambda\sum|w_i|$

$= n\sum w_i^2 - 2\sum y^T\vec{x}_i w_i + \sum y_i^2 + \lambda\sum|w_i|$

$\Rightarrow \boxed{\min_{w_i}\left(nw_i^2 - 2(y^T\vec{x}_i)w_i + y_i^2 + \lambda|w_i|\right)}$

(b) if $w_i^* > 0$

derivative of objective function

$2nw_i - 2y^Tx_i + \lambda = 0$ set equal to zero

$\Rightarrow |w_i| = w_i$

(by definition of abs. value $|x| = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}$)

$\boxed{w_i^* = \frac{2y^Tx_i - \lambda}{2n}}$

$\boxed{y^Tx_i > \frac{\lambda}{2}}$

and $y^Tx_i > 0$

Scanned with CamScanner

* (c) $w_i^* < 0$

by definition of $|\cdot|$

derivative

$|w_i| = -w_i$

$$2n w_i - 2y^T x_i - \lambda = 0$$

$$\boxed{w_i^* = \frac{2y^T x_i + \lambda}{2n}}$$

$$\boxed{|y^T x_i| > \frac{\lambda}{2}}$$

and $y^T x_i < 0$

* (d) if $|y^T x_i| \leq \frac{\lambda}{2}$

if $y^T x_i > 0$

$\Rightarrow \quad y^T x_i \leq \frac{\lambda}{2}$ and

if $y^T x_i < 0$

$-y^T x_i \leq \frac{\lambda}{2}$

$y^T x_i \geq -\frac{\lambda}{2}$ and $\boxed{w_i^* < 0}$

(e) Ridge Regression

$$w^* = \arg\min_{w \in \mathbb{R}^d} \underbrace{\|Xw - y\|_2^2 + \lambda\|w\|_2^2}$$

objective function

$$w^T X^T X w - 2y^T X w + y^T y + \lambda w^T w$$

$$n w^T w - 2y^T X w + y^T y + \lambda w^T w$$

component-wise

$$(n+\lambda)\, w_i^2 - 2(y^T x_i)\, w_i + y_i^2$$

$\therefore$ the problem is equivalent to

$$\min_{w_i} (n+\lambda) w_i^2 - 2(y^T x_i) w_i + y_i^2$$

$$\Rightarrow \quad w_i^* = \frac{2y^T x_i}{2(n+\lambda)} = \frac{y^T x_i}{n+\lambda}$$

which removes restrictions on the previous condition on $y^T x_i$, which was dependent on the choice of λ. This seems to be a more more robust form of regression

2. Image Compression

full SVD

(a) $\underset{m\times n}{A} = U \Sigma V^T$

$$= \underbrace{\begin{bmatrix} u_1 \cdots u_r \; u_{r+1} \cdots u_m \end{bmatrix}}_{m\times m} \underbrace{\left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & & \cdot \end{array}\right]}_{m\times n} \underbrace{\begin{bmatrix} - \; v_1^T \; - \\ \vdots \\ - \; v_r^T \; - \\ - \; v_{r+1}^T \; - \\ \vdots \\ - \; v_n^T \; - \end{bmatrix}}_{n\times n}$$

$\Rightarrow$ rank k approximation

$$\boxed{\underbrace{\begin{bmatrix} u_1 \cdots u_k \end{bmatrix}}_{m\times k} \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix}}_{k\times k} \underbrace{\begin{bmatrix} - \; v_1^T \; - \\ \vdots \\ - \; v_k^T \; - \end{bmatrix}}_{k\times n} = A_{m\times n}}$$

(b) $\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\mathrm{trace}(A^T A)} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

$$\boxed{\min_{\hat{y}} \|A - \hat{y}\|_F + \lambda \|\hat{y}\|_1}$$

c. Jupyter

3. Image Restoration

(a) $F = \underbrace{\begin{pmatrix} f(1,1) & \cdots & f(1,W) \\ & \vdots & \\ f(H,1) & \cdots & f(H,W) \end{pmatrix}}_{\text{Width}} \Bigg\}$ Height

since $f(i,j)$ represent the grayscale value of the image at discrete (pixel) coordinate (i,j)

we can represent the image as a matrix as shown above.

So it makes sense $F(i,j) = f(i,j)$

(b) $F(i,j) = f(i,j)$

$$\Rightarrow \nabla F(i,j) = \nabla f(i,j) = \begin{pmatrix} \frac{f(i+h,j) - f(i,j)}{h} \\ \frac{f(i,j+h) - f(i,j)}{h} \end{pmatrix}$$

for $h = 1$

$$\boxed{\nabla F(i,j) = G(i,j) = \begin{pmatrix} f(i+1,j) - f(i,j) \\ f(i,j+1) - f(i,j) \end{pmatrix}}$$

(c)

$$\min_{\substack{\hat{f} \\ \forall (x,y) \notin A}} \iint_{\Omega} \|\nabla \hat{f}(x,y)\|_2 \, dx dy$$

expressed as discrete Reimann sum over pixel values

$$\boxed{\min_{\substack{\hat{F} \\ \forall (i,j) \notin A}} \sum_{i=1}^{H} \sum_{j=1}^{W} \|\nabla \hat{F}(i,j)\|_2 \underbrace{\Delta i}_{=1} \underbrace{\Delta j}_{=1}}$$

OR

$$= \min_{\substack{\hat{F} \\ \forall (i,j) \notin A}} \sum_{i=1}^{H} \sum_{j=1}^{W} \sqrt{\left(f(i+1,j) - f(i,j)\right)^2 + \left(f(i,j+1) - f(i,j)\right)^2}$$

(d)

$$\min_{\substack{\hat{F}, y \\ \forall i,j \notin A}} \sum_{i=1}^{H} \sum_{j=1}^{W} y_j$$

$$\text{s.t} \quad \|\nabla \hat{F}(i,j)\|_2 \leq y_j \quad j = 1, \dots, W$$

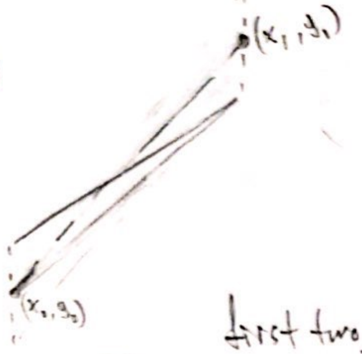
(e) In notebook.

It worked

Scanned with CamScanner

4. Slalom problem

(a)



first two points

$$\min \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

$$\min_{y.} \sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

$$\text{s.t. } y_i \geq y_i - c_i/2$$

$$y_i \leq y_i + c_i/2$$

as SOCP

$$\boxed{\begin{aligned} \min_{y,t} \ & \sum_{i=0}^{n-1} t_i \\ \text{s.t. } & \sqrt{(x_{i+1}-x_i)^2 + (y_{i+1}-y_i)^2} \leq t_i \\ & y_i \geq y_i - c_i/2 \\ & y_i \leq y_i + c_i/2 \end{aligned}}$$

(b) notebook

5. Sphere enclosure

$\min_{x,R} B(x,R)$

we want to find the minimum radius R of ball centered at x

subject to

$\|x - x_i\|_2 + \rho_i \leq R$

distance between center of B to center of each enclosed Ball B_i

→ this distance plus the radius ρ_i should be less than minimum radius R for each B_i

cast as SOCP

$$\boxed{\begin{aligned} &\min_{x,R} R \\ &\text{s.t. } B(x,r) \leq R \\ &\|x - x_i\|_2 + \rho_i \leq R \end{aligned}}$$