1.) a.) Define the objective function f for orthogonal regression

for $\underline{0} \neq \underline{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$

hyperplane is $H_{\underline{x},y} := \{ \underline{a} \in \mathbb{R}^n : \underline{x} \cdot \underline{a} = y \}$

for the data points $\underline{a}_1, \dots, \underline{a}_m$. In the orthogonal regression problem, want to find $\underline{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$:

$$\min_{\underline{x},y} \left\{ \sum_{i=1}^{m} d(\underline{a}_i, H_{\underline{x},y})^2 : \underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}, \; y \in \mathbb{R} \right\}$$

where $d(\underline{a}_i, H_{\underline{x},y})^2 = \dfrac{(\underline{a}_i \cdot \underline{x} - y)^2}{\|\underline{x}\|^2}$ for $i = 1, \dots, m$.

$$\therefore \min_{\underline{x},y} \left\{ \sum_{i=1}^{m} \frac{(\underline{a}_i \cdot \underline{x} - y)^2}{\|\underline{x}\|^2} : \underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}, \; y \in \mathbb{R} \right\}$$

So the objective function is $f(\underline{x}, y) = \sum_{i=1}^{m} \dfrac{(\underline{a}_i \cdot \underline{x} - y)^2}{\|\underline{x}\|^2}$

lecture notes 8 page 8

when $\underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}$, $y \in \mathbb{R}$

b.) If $H_{\underline{a},b}$ is the optimal hyperplane, write down a general formula for $\underline{a}$ and b.

hyperplane is $H_{\underline{a},b} := \{ \underline{x} \in \mathbb{R}^n : \underline{a}^T \underline{x} = b \}$

$\underline{x}_1, \underline{x}_2, \dots, \underline{x}_m$ are given data points. where $\underline{a}^n \in \mathbb{R}^n \setminus \{\underline{0}\}$, $b \in \mathbb{R}$.

$$\underline{a}^{k+1} = \underline{a}^{k} - t\nabla R_b(\underline{a}^{k})$$

fix $\underline{a}$ minimizing first first with respect to b we get

$$b = \frac{1}{m}\sum_{i=1}^{m} \underline{n}_i \cdot \underline{a} = \frac{1}{m}\underline{1}\cdot(\underline{A}\,\underline{a}) \qquad \underline{1} := \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} \in \mathbb{R}^n$$

$m = 10$ in this case.

$$\text{so } \sum_{i=1}^{m}(\underline{n}_i\cdot\underline{a} - b)^2 = \sum_{i=1}^{m}\left(\underline{n}_i\cdot\underline{a} - \frac{1}{m}\underline{1}\cdot(A\underline{a})\right)^2$$

$$= \sum_{i=1}^{m}(\underline{n}_i\cdot\underline{a})^2 - \frac{2}{m}\sum_{i=1}^{m}(\underline{n}_i\cdot\underline{a})\left(\underline{1}\cdot(A\underline{a})\right) + \frac{1}{m}\left(\underline{1}\cdot(A\underline{a})\right)^2$$

$$= \sum_{i=1}^{m}(\underline{n}_i\cdot\underline{a})^2 - \frac{1}{m}\left(\underline{1}\cdot(A\underline{a})\right)^2 = \|A\underline{a}\|^2 - \frac{1}{m}\left(\underline{1}\cdot(A\underline{a})\right)^2$$

$$= \underline{a}^T A^T\left(I_m - \frac{1}{m}\underline{1}\cdot\underline{1}^T\right)A\underline{x}$$

$$\text{so } \min_{\underline{a}}\left\{\frac{\underline{a}^T\left[A^T\left(I_m - \frac{1}{m}\underline{1}\cdot\underline{1}^T\right)A\right]\underline{x}}{\|\underline{a}\|^2} : \underline{a}\in\mathbb{R}^n\setminus\{\underline{0}\}\right\}$$

for symmetric matrix $A \in \mathbb{R}^{n\times n}$

The optimal solution of the orthogonal regression problem $(\underline{a}, b)$ is to take $\underline{a}$ to be the an eigenvector of associated with the minimum eigenvalue of $A^T(I_m - \frac{1}{m}\underline{1}\,\underline{1}^T)A$

as seen in lecture notes 8 page 4.

c.) Show that the Gradient method can be applied to the Rayleigh quotient to find the eigenvalue of the matrix in (b). In particular,

i) The constant stepsize can be applied

The Rayleigh quotient is

$$R_{\underline{A}}(\underline{a}) = \frac{\underline{a} \cdot (A\underline{a})}{\|\underline{a}\|^2} \qquad \forall\, \underline{a} \in \mathbb{R} \setminus \{\underline{0}\}$$

$$\|\underline{a}\|^2 = \sum_i a_i^2$$

$$R_A(\underline{a}) = \frac{1}{\|\underline{a}\|^2} \cdot \underline{a} \cdot (A\underline{a})$$

let $f(\underline{a}) = \underline{a}^T \cdot (A\underline{a})$

$g(\underline{a}) = \underline{a}^T \cdot \underline{a}$

$$R_A(\underline{a}) = \frac{f(\underline{a})}{g(\underline{a})}$$

so $$\nabla R_A(\underline{a}) = \frac{f'(\underline{a})\, g(\underline{a}) - f(\underline{a})\, g'(\underline{a})}{(g(\underline{a}))^2}$$

$$\left.\begin{aligned} f'(\underline{a}) &= A\underline{a} + \underline{a}^T A = 2A\underline{a} \\ g'(\underline{a}) &= \underline{a} + \underline{a}^T = 2\underline{a} \end{aligned}\right\} \Rightarrow \begin{aligned} f'(\underline{a})\, g(\underline{a}) &= (2A\underline{a})\, \underline{a}^T\underline{a} \\ f(\underline{a})\, g'(\underline{a}) &= (\underline{a}^T A \underline{a})(2\underline{a}) \end{aligned}$$

so $f'(\underline{a})\, g(\underline{a}) - f(\underline{a})\, g(\underline{a}) = 2\left((A\underline{a})\|\underline{a}\|^2 - (\underline{a}^T A \underline{a}) \cdot \underline{a} \right)$

↳ numerator is now $-2\left((A\underline{a})\|\underline{a}\|^2 - (\underline{a}^T A \underline{a})\underline{a}\right)$

↳ denomenator is now — $\|a\|^2 \cdot \|a\|^2$

$$\nabla R_A(a) = 2\frac{(Aa)\|a\|^2 - (a^T A a)a}{\|a\|^2 \cdot \|a\|^2} = 2\frac{Aa}{\|a\|^2} - 2\frac{R_A(a)\cdot a}{\|a\|^2}$$

$$\nabla R_A(a) = 2\left(\frac{Aa - R_A(a)\cdot a}{\|a\|^2}\right)$$

Since $a \in \mathbb{R}^n \setminus \{0\} \Rightarrow \nabla R_A(a)$ is well defined.

$\Downarrow$

with a constant step size being des applied.

There can be investigation done to find the eigenvalue of the matrix. as the $\exists\, \nabla R_A(a)$ means there should be convergence.

ii.) if x^k is an iteration point, then $x^{k+1} \neq 0$

$$x^{k+1} = x^k - t\nabla R_A(x^k) \qquad \text{because constant stepsize.}$$

$$\|x^{k+1}\|^2 = \|x^k - t\nabla R_A(x^k)\|^2 = (x^k - t\nabla R_A(x^k))\cdot(x^k - t\nabla R_A(x^k))$$

$$= \|x^k\|^2 - t\, x^k\cdot\nabla R_A(x^k) - t\nabla R_A(x^k)\cdot x^k + t^2\|\nabla R_A(x^k)\|^2$$

since $\nabla R_A(x^k) = 2\left(\dfrac{Ax^k - R_A(x^k)\cdot(x^k)}{\|x^k\|^2}\right) \in \mathbb{R}^n$

$$\|\underline{x}^{k+1}\|^2 = \|\underline{x}^k\|^2 - 2t\,\underline{\nabla} R_A(\underline{x}^k) + t^2 \|\underline{\nabla} R_A(\underline{x}^k)\|^2$$

Since $\underline{x}^k \in \mathbb{R}^n \setminus \{\underline{0}\} \Rightarrow \|\underline{x}^k\|^2 > 0$

$$\|\underline{x}^{k+1}\|^2 > -2\underline{x}^k \cdot \underline{\nabla} R_A(\underline{x}^k) + t^2 \|\underline{\nabla} R_A(\underline{x}^k)\|^2$$

and $\because \|\underline{\nabla} R_A(\underline{x}^k)\|^2 \geq 0$

$$\|\underline{x}^{k+1}\|^2 > -2\underline{x}^k \cdot \underline{\nabla} R_A(\underline{x}^k) = -4\,\underline{x}^k \left(\frac{A\underline{x}^k - R_A(\underline{x}^k) \cdot \underline{x}^k}{\|\underline{x}^k\|^2} \right)$$

$$\|\underline{x}^{k+1}\|^2 > -4 \left(\frac{(\underline{x}^k)^T A \underline{x}^k}{\|\underline{x}^k\|^2} - \frac{(\underline{x}^k)^T R_A(\underline{x}^k) \cdot \underline{x}^k}{\|\underline{x}^k\|^2} \right)$$

$$> -4 \left(R_A(\underline{x}^k) \left(1 - \frac{\|\underline{x}^k\|^2}{\|\underline{x}^k\|^2} \right) \right) \quad \text{Since } R_A(\underline{x}^k) \in \mathbb{R}$$

$$> -4 \left(R_A(\underline{x}^k) \times 0 \right) = 0$$

Since $\|\underline{x}^{k+1}\|^2 > 0 \quad \therefore \quad \underline{x}^{k+1} \neq 0$.

d.) Using the gradient method on the Rayleigh quotient, provide a point $(\underline{x}, y)$: $0.45 < f(\underline{x}, y) < 1.54$

So $\underline{x}^{k+1} = \underline{x}^k - t\,\nabla R_A(\underline{x}^k)$ we implement the gradient descent method to iteratively update $\underline{x}$ and y.

Until the objective function $f(\underline{x}, y)$ is in the interval

Implementing this in

Implementing this in python I got the optimal values for $\underline{x}, y$ and $f(\underline{x}, y)$ to be

$$\underline{x}^* = [-0.507, -0.214, -0.824]^T \qquad y^* = -2.281$$

$$f(\underline{x}^*, y^*) = 0.453 \qquad \text{all to 3 decimal place.}$$

2.) As C.W. moves to london she needs to find accommodation. She is very busy:

i.) She goes to her collage four days a week at $a_1 = (1,2)$

ii.) On Monday she plays chess at a friend's house $a_2 = (3,0)$

iii.) on Tuesday and Thursday she plays in a badminton tournament at $a_3 = (3,1)$

iv.) on Friday, Saterday and Sunday she plays in football tournament at $a_4 = (2,3)$.

C.W. loves walking, she studing at the university library, and almost does not spend a single minute at home. Then, she seeks an accomidation which minimises the cost of the subway. You can also assume that

v.) the cost of transportation is one pound for each kilometer.

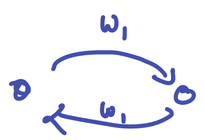
vi.) before going to any other place she needs first to come back home.

Prove that the Fermat-Weber problem is suitable for this problem. In particular:

a.) Explain which quantities are the anchors and the weights.

The anchors are positions which C.W. is required to visit frequently, which are a_1, a_2, a_3, a_4 which are collage, friends house (chess), Badminton court and football pitch respectively.

The weight would be how frequently C.W visits each location such as the weight for a_1 would be 8. Due to condition (vi) where if she goes to a_1 then if she wishes to go to a_2 then she must return home. Hence it is doubled.



$\Rightarrow$ $2w_1$ = actual weight for w_1

round trip.

So the actual weights are

$w_1 = 8$	for	$a_1 = (1,2)$	- collage
$w_2 = 2$	for	$a_2 = (3,0)$	- friends house (chess)
$w_3 = 4$	for	$a_3 = (3,1)$	- badminton tourniment
$w_4 = 6$	for	$a_4 = (2,3)$	- football tourniment

This is switable for Fermat-Weber problem since we have a weight - which is how frequently a location is visited and we have a location.

The fermat-Weber Problem where $w_i > 0$ $\underline{a}_i \in \mathbb{R}^n$ find $\underline{x} \in \mathbb{R}^n$ to minimise the weighted distance of $\underline{x}$ to each w_i

$$\min_{\underline{x} \in \mathbb{R}^n} \left\{ f(\underline{x}) := \sum_{i=1}^{m} w_i \| \underline{x} - \underline{a}_i \| \right\}$$

The objective fucntion are not differentiable at the anchor points $\underline{a}_i$

b.) Prove that the Weiszfeld method can be applied to solve this problem and apply it. In particular:

i.) show that none of the anchors is a minimum

So Weiszfeld method assume $\underline{x}$ optimum is not an anchor point

so $\underline{x}^k \neq \underline{a}_i$ the objective function $\underline{x}^{k+1} = \underline{x}^k - t^k \underline{\nabla} f(\underline{x}^k)$

The first order optimality condition mentioned in Lecture notes 5 page 3.

$\|\cdot\|_2 := \|\cdot\|$

$\underline{\nabla} f(\underline{x}^k) = 0$. using the fact $\nabla \|\underline{x}^k\| = \dfrac{\underline{x}^k}{\|\underline{x}^k\|}$

$$\text{so} \quad \underline{\nabla} f(\underline{x}^k) = \sum_{i=1}^{m} \frac{w_i(\underline{x}^k - \underline{a}_i)}{\|\underline{x}^k - \underline{a}_i\|}$$

$$\text{and} \quad t^k = \frac{1}{\displaystyle\sum_{i=1}^{m} \frac{w_i}{\|\underline{x}^k - \underline{a}_i\|}} \qquad \lambda_i = \sum_{i=1}^{m} \left[\frac{\dfrac{w_i}{\|\underline{x}^k - \underline{a}_i\|}(\underline{x}^k - \underline{a}_i)}{\displaystyle\sum_{i=1}^{m} \frac{w_i}{\|\underline{x}^k - \underline{a}_i\|}} \right]$$

$$\text{so} \quad t^k \underline{\nabla} f(\underline{x}^k) = \sum_{i=1}^{m} \lambda_i (\underline{x}^k - \underline{a}_i)$$

$$= \lambda_1(\underline{x}^k - \underline{a}_1) + \lambda_2(\underline{x}^k - \underline{a}_2) + \dots + \lambda_m(\underline{x}^k - \underline{a}_m)$$

Since the simplex $\lambda_i \in (0,1) \quad \forall i = \{1, 2, \dots, m\}$

$\therefore \quad \sum_{i=1}^{m} \lambda_i = 1$ Since λ_i is open ball $t^k \nabla f(\underline{x}^k)$ will get close but never equal to $(\underline{x}^k - \underline{a}_i)$

so $\underline{x}^{k+1} = \underline{x}^k - t^k \nabla f(\underline{x}^k) \neq \underline{x}^k - (\underline{x}^k - \underline{a}_i)$

$\underline{x}^{k+1} \neq \underline{a}_i$ because next iteration is never $\underline{a}_i$

$\therefore$ the iteration $\underline{x}^{k+1}$ never reaches the anchor point hence non of the anchors is a minimum.

ii.) given an iteration point $\underline{x}^k$, it is possible to define $\underline{x}^{k+1}$

The stationary condition can be written as a fixed point $\underline{x} = T(\underline{x})$

where $T(\underline{x}) = \dfrac{1}{\sum_{i=1}^{m} \frac{w_i}{\|\underline{x} - \underline{a}_i\|}} \sum_{i=1}^{m} \dfrac{w_i \cdot \underline{a}_i}{\|\underline{x} - \underline{a}_i\|}$ To find the optimal $\underline{x}$ is finding the fixed point.

The Weiszfield proposed a fixed point iteration method:

$$\underline{x}^{k+1} = T(\underline{x}^k)$$

$$\underline{x}^{k+1} = \underline{x}^k - \frac{1}{\sum_{i=1}^{m} \frac{w_i}{\|\underline{x}^k - \underline{a}_i\|}} \nabla f(\underline{x}^k).$$

with the stepsize is

$$t^k = \frac{1}{\sum_{i=1}^{m} \frac{w_i}{\|\underline{x}^k - \underline{a}_i\|}}$$

from part i.) we know t^k and $\nabla f(\underline{x}^k)$ are well defind.

c.) Keep the weekly transport expenses below £22.74

This is not possiable. As when the performing the iteration above. The Weiszfeld method.

$$\underline{\nabla} f(\underline{x}^k) = \sum_{i=1}^{m} \frac{w_i \cdot \underline{a}_i}{\| \underline{x} - \underline{a}_i \|} \quad \text{for } \| \underline{\nabla} f(\underline{x}^k) \| < 10^{-6}$$

$f : \mathbb{R}^2 \to \mathbb{R}$ So the hessian denote as

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} \quad \text{where } H \in \mathbb{R}^{2\times2}$$

for the given point $H(\underline{x}^{172}) > 0$ positive definite.

This proves the appartment cordinate is optimal as it is bellow the tolerance value 10^{-6} for the gradint. Further more the hessian is positive definite with the given weights and cordinats. This means

$\underline{x}^* = [1\,377, 2\,059]$ is the local minimum. in metere

at $k = 172$

With an initial starting point of $\underline{x}^0 = [0, 0]$

This then comes down to the weekly transportation expenses to be £ 22.82 if living in the $\underline{x}^*$ position.

The wealy transportation expenses are $\sum_{i=1}^{m} w_i \| \underline{x}^* - \underline{a}_i \| \times \text{cost}$

cost $= £1$ since it is £1 per kilometer.