การบ้าน HW1 ของ
นาย ธนิต ตติวรรณรัตน์ 6031024921

ปรึกษา การทำการบ้านกับ
นายพสิษฐ์ เสาหะวรุตม์ชัย
6031039321

(T1) กำหนดให้ $A_{m \times n}$, $B_{n \times m}$, $Y = AB$

จะได้ $Y_{i,j} = \sum_x A_{i,x} B_{x,j}$

Y จะมี dimension เป็น $m \times m$

ดังนั้น $Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ \vdots & \ddots & & \vdots \\ Y_{m1} & & \cdots & Y_{mm} \end{bmatrix}$

$$Y = \begin{bmatrix} \sum_x A_{1,x} B_{x,1} & \cdots & \sum_x A_{1,x} B_{x,m} \\ \vdots & \sum_x A_{2,x} B_{x,2} & \\ & & \ddots \\ \sum_x A_{m,x} B_{x,1} & \cdots & \sum_x A_{m,x} B_{x,m} \end{bmatrix}$$

$$tr(Y) = tr\, AB$$

$$= \sum_y \sum_x A_{y,x} B_{x,y}$$

ดังนั้น $\nabla_A f(A)$ ณ ตำแหน่ง (i,j)

จะได้ว่า

$$\frac{\partial}{\partial A_{i,j}} \left(\sum_y \sum_x A_{y,x} B_{x,y} \right) = \frac{\partial}{\partial A_{i,j}} \left(\sum_x A_{1,x} B_{x,1} + \cdots + \sum_x A_{i,x} B_{x,i} + \cdots + \sum_x A_{m,x} B_{x,m} \right)$$

(terms other than $\sum_x A_{i,x} B_{x,i}$ crossed out → 0)

$$= \frac{\partial}{\partial A_{i,j}} \left(\sum_x A_{i,x} B_{x,i} \right)$$

$$= \frac{\partial}{\partial A_{i,j}} \left(A_{i,1} B_{1,i} + \cdots + A_{i,j} B_{j,i} + \cdots + A_{i,m} B_{m,i} \right)$$

(terms other than $A_{i,j} B_{j,i}$ crossed out → 0)

$$= \frac{\partial}{\partial A_{i,j}} \left(A_{i,j} B_{j,i} \right)$$

$$= B_{j,i} = B_{i,j}^T$$

ตัวที่ตำแหน่ง (i,j) ของ $\nabla_A tr(AB)$ คือ $B_{j,i} = B_{i,j}^T$

$$\therefore \nabla_A tr(AB) = B^T$$

(2) กำหนดให้ A เป็น matrix ขนาด $m \times n$

และ $\nabla_{A^T} f(A)$ ที่ตำแหน่ง (i,j)

คือ $\dfrac{\partial f(A)}{\partial A_{i,j}^T} = \dfrac{\partial f(A)}{\partial_{j,i}}$

$$\nabla_{A^T} f(A) = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{1,1}} & \cdots & \frac{\partial f(A)}{\partial A_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{n,1}} & \cdots & \frac{\partial f(A)}{\partial A_{n,m}} \end{bmatrix}_{n \times m}$$

$$\text{let } \nabla_A f(A) = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{1,1}} & \cdots & \frac{\partial f(A)}{\partial A_{1,n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m,1}} & \cdots & \frac{\partial f(A)}{\partial A_{m,n}} \end{bmatrix}_{m \times n}$$

$$\text{and } \left(\nabla_A f(A)\right)^T = \begin{bmatrix} \frac{\partial f(A)}{\partial A_{1,1}} & \cdots & \frac{\partial f(A)}{\partial A_{1,m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{n,1}} & \cdots & \frac{\partial f(A)}{\partial A_{n,m}} \end{bmatrix}_{n \times m}$$

$$\therefore \nabla_{A^T} f(A) = \left(\nabla_A f(A)\right)^T$$

(13) $\nabla_A \operatorname{tr}(ABA^TC) = \nabla_A \operatorname{tr}((AB)(A^TC))$

$= \nabla_{A:AB} \operatorname{tr}((AB)(A^TC)) + \nabla_{A:A^TC} \operatorname{tr}((AB)(A^TC))$

$= (A^TC)^T \nabla_A \operatorname{tr}(AB) + \left(\nabla_{A^T:A^TC} \operatorname{tr}((AB)(A^TC))\right)^T$

$= C(A^T)^T B^T + \left((AB)^T \nabla_{A^T}(A^TC)\right)^T$

$= C^TAB^T + \left((AB)^T C^T\right)^T$

$= C^TAB^T + CAB$

$\therefore \nabla_A \operatorname{tr}(ABA^TC) = CAB + C^TAB^T$

공식 $\nabla_A(AB) = B^T$

공식 $\nabla_{A^T} f(A) = (\nabla_A f(A))^T$

공식 $\nabla_A \operatorname{tr}(ABI) = \nabla_A \operatorname{tr}(BIA)$

$= \nabla_A \operatorname{tr}(IAB)$

$\therefore \nabla_A \operatorname{tr}(AB) = \nabla_A \operatorname{tr}(BA)$

$= B^T$

(14) ข้อมูลเรียงตามนี้ $[(3,3),(2,2),(-3,-3),(8,8),(6,6),(7,7),(-2,-4),(-7,-7),(1,2)]$

ตอนแรก ให้ $(3,3),(2,2),(-3,-3)$ เป็น ตัวแทน ของ centroid แต่ละ cluster

ต่อมา นำค่า $(8,8)$ เพื่อมาหา distance ระหว่าง $(8,8)$ กับ centroid แต่ละ cluster แล้ว นำ $(8,8)$ เข้าไป เป็น สมาชิก ของ cluster ที่ distance น้อยที่สุด ซึ่งก็คือ cluster ที่ 1 แล้วทำการ update centroid โดย หาค่าเฉลี่ย ของ x และ y เพื่อเป็น centroid ใหม่ จะได้ list ของ centroid เป็น $(5.5, 5.5),(2,2),(-3,-3)$ และมี สมาชิก cluster เป็น $[(3,3),(8,8)],[(2,2)],[(-3,-3)]$ ต่อมา นำ $(6,6)$ เข้าไปหา distance กับ centroid แล้ว นำ $(6,6)$ ไป เป็น สมาชิก cluster ที่ centroid ใกล้ที่สุด จะ update centroid เป็น $(5.7, 5.7),(2,2),(-3,-3)$ และ สมาชิก cluster เป็น $[(3,3),(8,8),(6,6)],[(2,2)],[(-3,-3)]$ ต่อมา นำ $(7,7)$ ไปหา centroid ที่ ใกล้ที่สุด แล้วนำ เข้า cluster จะได้ $[(3,3),(8,8),(6,6),(7,7)],[(2,2)],[(-3,-3)]$ และ update centroid ได้เป็น $(6,6),(2,2),(-3,-3)$ แล้วนำ $(-2,-4)$ ไปหา centroid ที่ ใกล้ที่สุด แล้วนำ เข้า cluster ของ centroid นั้น โดย แต่ละ cluster จะเป็น $[(3,3),(8,8),(6,6),(7,7)],[(-2,-2)],[(-3,-3),(-2,-4)]$ แล้วทำการ update centroid จะได้ $(6,6),(2,2),(-2.5,-3.5)$ ต่อมา นำ $(-7,-7)$ มาหา centroid ที่ ใกล้ที่สุด แล้วนำ เข้า cluster โดยจะมี ข้อมูลเป็น $[(3,3),(8,8),(6,6),(7,7)],[(2,2)],[(-3,-3),(-2,-4),(-7,-7)]$ ต่อมา จะนำ $(1,2)$ มาหา distance น้อยที่สุด กับ centroid แล้ว บรรจุ เข้า cluster ของ centroid นั้น ซึ่งจะได้เป็น

$\therefore [(3,3),(8,8),(6,6),(7,7)],[(2,2),(1,2)],[(-3,-3),(-2,-4),(-7,-7)]$

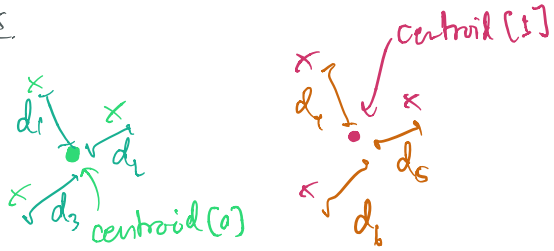
และมี centroid เป็น $(6,6),(1.5,2),(-2.5,-3.5)$

(15) จะได้เป็น $[(-3,-3),(-2,-4)],[(2,2),(1,2),(8,8),(6,6),(7,7),(3,3)],[(-7,-7)]$

และมี centroid เป็น $(-2.5,-3.5),(4.5,4.7),(-7,7)$ ตามลำดับ

(T6) ให้หา distance ของแต่ละ centroid กับ จุดใน cluster ที่ centroid ตัวเองตัว represent อยู่ แล้วนำมาบวกกัน

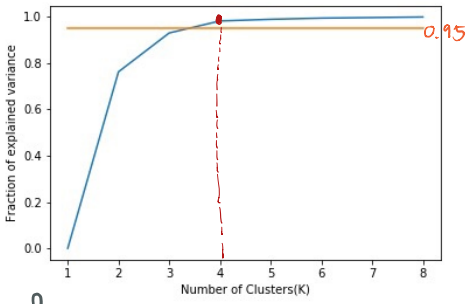
ex.



ในกรณีที่ k=2 จะได้ loss = $d_1 + d_2 + \dots + d_6$

(T7) k ที่ดีที่สุด คือ k = 4

เพราะเป็น k ที่น้อยที่สุดสามารถ represent ได้อย่างน้อย 95% ของ all-variance data



(T7) median ของ age ใน training set เท่ากับ 28.0

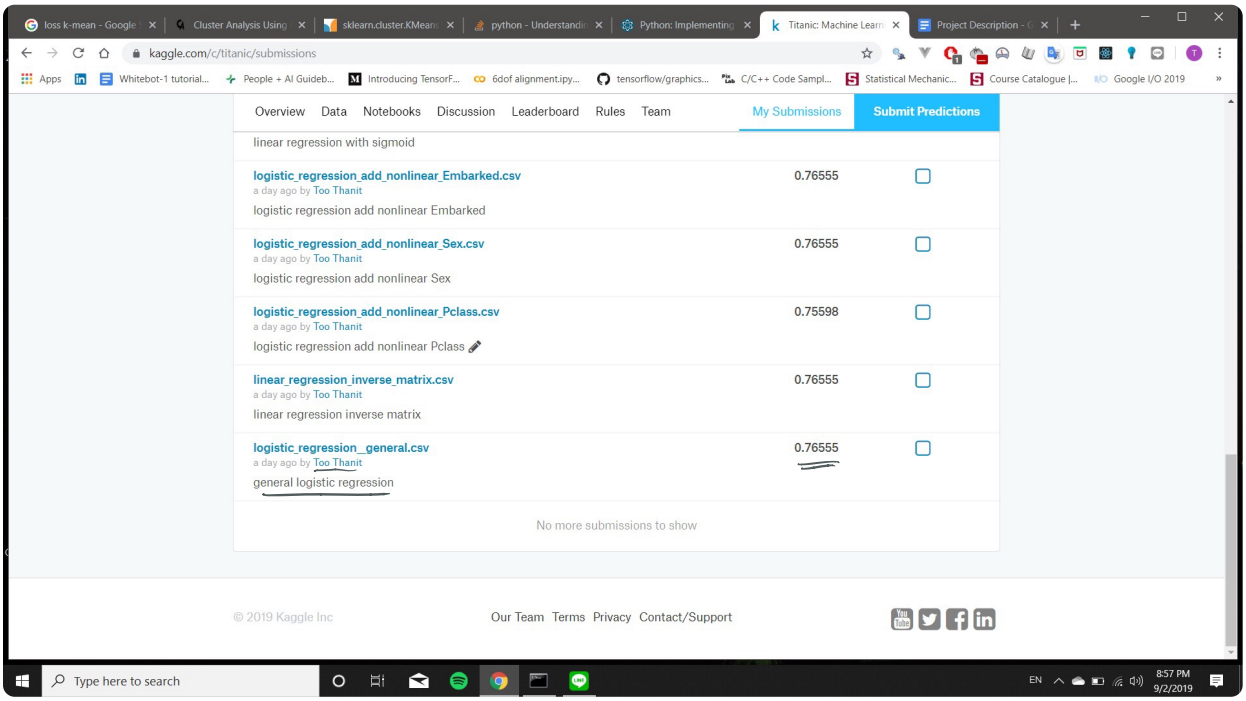
(T8) mode ของ embarked ใน training set เป็น 0

(T9) มี code แบบใช้แล้ว ที่ไฟล์ logistic_regression_general

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

$$\theta_j \leftarrow \theta_j + lr\left(\sum_i (y_i - \theta^T x_i)\, x_i^{(j)}\right)$$

T 10



gradient descent approach

OT2 แนบ code ที่ใช้ทำไฟล์ linear_regression_with_sigmoid.

และไฟล์ linear_regression_inverse_matrix

↑ inverse matrix approach

OT3 หาค่า MSE

- linear_regression_with_sigmoid ได้ MSE = 1.03897
- linear_regression_inverse_matrix ได้ MSE = 1.03885

MSE ของทั้ง 2 วิธีที่คำนวณได้ต่างกันเล็กน้อย

(OT4) test_score ลดลงนิดหน่อย เมื่อเพิ่มทีหลังสองของ feature "P class"

สามารถดู code ได้ที่ logistic_regression_add_nonlinear

(OT5) test_score ลดลงนิดหน่อย เมื่อเหลือแค่ 2 features

คือ Sex กับ Age สามารถดู code ได้ที่ logistic_regression_2_feature_Sex_Age